Working Paper 98-18
Statistics and Econometrics Series 11
February 1998

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

NON-GAUSSIAN LOG-PERIODOGRAM REGRESSION.

Carlos Velasco*

Abstract

We show the consistency of the log-periodogram estimate of the long memory parameter for long range dependent linear, non necessarily Gaussian, time series when we make a pooling of periodogram ordinates. Then, we study the asymptotic behaviour of the tapered periodogram of long range dependent time series for frequencies near the origin. Finally, we obtain the asymptotic distribution of the log-periodogram estimate for possibly non-Gaussian observations when we use the tapered periodogram. For that result we rely on higher order asymptotic properties of a vector of periodogram ordinates of the linear innovations.

Keywords:
Long range dependence, semiparametric inference, higher order asymptotics, tapering.

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. C/ Madrid, 126 28903 Madrid. Spain. Ph: 34-1-624.98.87, Fax: 34-1-624.98.49, e-mail: cavelas@est-econ.uc3m.es. The first version of this paper was written while the author was at the London School of Economics and Political Science and is based on the author´s Ph.D. Thesis. I thank P.M. Robinson and F.J. Hidalgo for helpful discussions and suggestions. Financial support from the Fundación Ramón Areces (Spain) and the Economic and Social Research Council (ESRC) grant n. R000235892 is gratefully acknowledged. Research funded by the Spanish Dirección General de Enseñanza Superior, Ref. n. PB95-0292.

Non-Gaussian Log-Periodogram Regression*

Carlos Velasco

Department of Statistics and Econometrics

Universidad Carlos III de Madrid

C. Madrid 126

28903 Getafe (Madrid)

Spain Tel. +34 1 6249887

February 2, 1998

*The first version of this paper was written while the author was at the London School of Economics & Political Science and is based on the author's PhD Thesis. I thank P.M. Robinson and F.J. Hidalgo for helpful discussions and suggestions. Financial support from the Fundación Ramón Areces (Spain) and the Economic and Social Research Council (ESRC) grant no. R000235892 is gratefully acknowledged.

Running head:

Non-Gaussian Log-Periodogram

Author and mailing address:

Carlos Velasco

Department of Statistics and Econometrics

Universidad Carlos III de Madrid

Calle Madrid 126

28903 Getafe (Madrid) Spain

Abstract

We show the consistency of the log-periodogram estimate of the long memory parameter for long range dependent linear, non necessarily Gaussian, time series when we make a pooling of periodogram ordinates. Then, we study the asymptotic behaviour of the tapered periodogram of long range dependent time series for frequencies near the origin. Finally, we obtain the asymptotic distribution of the log-periodogram estimate for possibly non-Gaussian observations when we use the tapered periodogram. For that result we rely on higher order asymptotic properties of a vector of periodogram ordinates of the linear innovations.

Key words: Long range dependence; semiparametric inference; higher order asymptotics; tapering.

1 Introduction

Long memory or long range dependent observations have been found in many fields of research (e.g. Robinson (1994c), Beran (1994)). In this paper we consider semiparametric statistical inference for long range stationary dependent time series. In particular we concentrate on the semiparametric estimate of the memory parameter based on the regression on the logarithm of the periodogram at Fourier frequencies close to the origin. This estimate, proposed initially by Geweke and Porter-Hudak (1983), has been very popular among practitioners because of its intuitive and computational appeal. However, properties of maximum likelihood methods have been analyzed extensively for parametric models of long range dependence (see, for example, Fox and Taqqu (1986) and Dahlhaus (1989)), obtaining equivalent efficiency results to the weak dependence situation. This approach involves a complete specification of the dynamics of the process. If we are only interested in the estimation of long range dependence characteristics, semiparametric and nonparametric set-ups are robust against any misspecification of the short run behaviour of the time series.

Semiparametric models for long memory focus on some properties of the autocovariance sequence (hyperbolic decay) or of the spectral density (singularity at the zero frequency). They are semiparametric because they do not make explicit assumptions on the behaviour of the autocovariances at short lags or on the spectral density apart from the origin.

We set our conditions in the frequency domain in terms of the spectral density since they are much more neat and cover a broader range of possibilities. We will assume that the spectral density satisfies

$$f(\lambda) \sim C\lambda^{-2d}, \qquad \text{as } \lambda \to 0^+, \tag{1}$$

where $d \in (0, \frac{1}{2})$ is the self similar parameter that governs the degree of strong dependence of the series. This is the interval of values of d for which the series exhibits long range dependence and is stationary. The basis for the log-periodogram estimate is the linear relationship implicit in expression (1) between the spectral density and the frequency in log-log coordinates with slope $-2d$

Robinson (1994a, 1995a and 1995b) and Lobato and Robinson (1996) have used similar assumptions to the ones we employ here to study the asymptotic behaviour of several semiparametric estimates of d. Robinson (1995a) justified a modified version of the procedure proposed by Geweke and Porter-Hudak (1983), including multivariate and pooled periodogram versions. He proved the consistency and asymptotic normality of this estimate for Gaussian vector time series. In this paper we extend his consistency results for linear processes not necessarily Gaussian. To obtain an asymptotically normal estimate we need to taper the data to reduce the leakage in the periodogram ordinates from the zero frequency pole and we need to pool the contribution for several adjacent frequencies in order to obtain a better behaved regressors.

The paper is organized as follows. In the next section we present our main assumptions and definitions and discuss related references. In Section 3 we obtain the consistency of the log-periodogram estimate of d. The effects of tapering are discussed in Section 4. Section 5 is dedicated to the asymptotic normality of the estimate of d when we used the tapered periodogram. Finally, we report the results of a brief simulation exercise centred on the tapering and pooling techniques analyzed. All the proofs are given in several appendices at the end of the paper.

2 Assumptions and definitions

Let $\{X_t,\ t = 1, 2, \ldots\}$ be a covariance stationary process with spectral density satisfying (1). Given an observable sequence X_t, $t = 1, \ldots, N$, we introduce the discrete Fourier transform at the frequency $\lambda_j = 2\pi j/N$, j integer,

$$w(\lambda_j) = (2\pi N)^{-1/2} \sum_{t=1}^{N} X_t e^{it\lambda_j},$$

and the periodogram

$$I(\lambda_j) = |w(\lambda_j)|^2 .$$

Define for $J = 1, 2, \ldots$, fixed, and some positive integers ℓ and m (assuming $(m - \ell)/J$ integer),

$$Y_k^{(J)} = \log\left(\sum_{j=1}^{J} I(\lambda_{k+j-J})\right) \qquad k = \ell + J, \ell + 2J, \ldots, m.$$

The estimate considered in Robinson (1995a) is

$$\widehat{d} = \left(\sum_k \Lambda_k^2\right)^{-1} \left(\sum_k \Lambda_k Y_k^{(J)}\right),$$

where $\Lambda_k = z_k - \overline{z}$, $\overline{z} = \{J/(m-\ell)\} \sum_k z_k$ and $z_k = -2 \log \lambda_k$. Here m is an integer smaller than N and ℓ is a user-chosen trimming number. In the asymptotics both numbers tend to infinity with the sample size N, but more slowly. We suppress in the notation reference to N or J.

We could substitute the (pooled) periodogram by non-parametric smoothed consistent estimates of the spectral density as was done in Velasco (1997). However, when we consider in $\widehat{d}$ fixed averages of the periodogram the analysis is much more complicated than in that situation. Here, we have to deal with the logarithm of a random variable which is not converging asymptotically to any constant and which can take values arbitrarily close to zero. Non-linear functions (the logarithm in particular) of the periodogram of stationary sequences have been considered under different set-ups (see e.g. Hannan and Nicholls (1977), Taniguchi (1979), Chen and Hannan (1980), von Sachs (1994a), Janas and von Sachs (1993), Robinson (1995a) and Comte and Hardouin, (1995a and 1995b), and the references given there). These works assume Gaussianity to obtain the main results, except Chen and Hannan, and Janas and von Sachs, who work with a linear process condition.

These last two references use higher order properties of the asymptotic distribution of the periodogram. Janas and von Sachs mainly apply the results for weakly dependent sequences of Götze and

Hipp (1983), making almost impossible to relax their assumptions for long range dependence situations. Instead, the approach of Chen and Hannan (1980) is based in the factorization of the periodogram of the observable sequence in the transfer function of the linear filter, times the periodogram of the independent and identically distributed (i.i.d.) innovations, plus a stochastic error term. The magnitude of this error depends on the smoothness of the spectral density and on the number of moments assumed for the innovations. Obviously the conditions they assumed, ($\sum |j|^{\delta}|\alpha_j| < \infty$, $\delta > \frac{1}{2}$, see Assumption 3 below), rule out any long memory behaviour or any singularity in the spectral density of X_t, but their results are based mainly on the properties of the periodogram of the i.i.d. innovations sequence, for which we assume the same set of conditions as in their Theorem 2 (see Assumption 4 below).

A related approach is used by Comte and Hardouin in a long-memory environment but assuming Gaussianity. We use one idea of them to avoid a modification of the estimate of d in the same spirit as the one Chen and Hannan (1980) proposed for a different statistic to account for *too small* values of $I(\lambda_j)/f(\lambda_j)$. Here, instead of redefining the periodogram with a truncation, we use an average of periodogram ordinates. Then we can use their higher order asymptotic approach and the long range dependence results of Robinson (1995b) to approximate the periodogram of X_t by that of the linear i.i.d. innovations times the long memory transfer function.

Tukey (1967) proposed tapering as an effective bias reduction technique for spectral inference to avoid leakage from remote frequencies. Under additional smoothness conditions on the behaviour of the spectral density at the origin, we study the asymptotic effect of tapering the data previously to calculate the periodogram. We obtain the asymptotic normality of the estimate $\widehat{d}$ based on the tapered periodogram. Von Sachs (1994a) and Janas and von Sachs (1993) used also tapering for non linear functions of the periodogram, but their results do not apply to long memory time series. Robinson (1986), Dahlhaus (1988) and Hurvich and Ray (1995), among others, have proposed this technique to reduce the bias of several statistics when a possible nonstationary behaviour of the observed time series is suspected.

Now we introduce some assumptions about the behaviour of the spectral density around the origin, following Robinson (1995a and 1995b), but not considering negative values of d. Later we will strengthen these assumptions to obtain further results.

Assumption 1 *X_t is covariance stationary and as $\lambda \to 0^+$,*

$$f(\lambda) = G\lambda^{-2d} + O(\lambda^{\alpha-2d}),$$

with $d \in [0, \frac{1}{2})$, $\alpha \in (0, 2]$ and $0 < G < \infty$.

Assumption 2 *In a neighbourhood $(0, \varepsilon)$ of the origin, $f(\lambda)$ is differentiable and*

$$\frac{d}{d\lambda} \log f(\lambda) = O(\lambda^{-1}), \qquad \text{as } \lambda \to 0^+.$$

These conditions are standard in long memory research and are satisfied with $\alpha = 2$ by fractional ARIMA models, for which

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left|1 - e^{i\lambda}\right|^{-2d} \left|\frac{a(e^{i\lambda})}{b(e^{i\lambda})}\right|^2, \qquad -\pi < \lambda \leq \pi,$$

where $\sigma^2 > 0$ and a and b are polynomials of finite degree having no zeros in or on the unit circle, and by the fractional noise model with autocovariance sequence given by

$$\gamma(j) = \frac{\gamma(0)}{2} \left(|j+1|^{2d+1} - 2|j|^{2d+1} + |j-1|^{2d+1}\right), \qquad j = \pm 1, \pm 2, \ldots$$

Instead of Gaussianity we introduce a fourth order stationary linear process condition, with filter coefficients compatible with Assumptions 1 and 2:

Assumption 3 *X_t satisfies*

$$X_t = \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j}, \qquad \sum_{j=0}^{\infty} \alpha_j^2 < \infty,$$

where $E[\epsilon_t] = 0$, $E[\epsilon_t^2] = 1$ and $E[\epsilon_t^4] < \infty$.

We assume zero mean for the series X_t without loss of generality, since we are omitting the periodogram at zero frequency in the definition of $\widehat{d}$. Four bounded moments are enough for all our consistency results. We introduce the next assumption following Chen and Hannan (1980):

Assumption 4 *The ϵ_t in Assumption 3 are i.i.d., with characteristic function $\widehat{Q}(\theta) = E[e^{i\theta\epsilon_t}]$ satisfying*

$$\sup_{|\theta| \geq \theta_0} \left|\widehat{Q}(\theta)\right| = \delta(\theta_0) < 1, \quad \forall \theta_0 > 0, \quad \textit{and}$$

$$\int_{-\infty}^{\infty} |\widehat{Q}(\theta)|^p d\theta < \infty, \qquad \textit{for some integer } \; p > 1.$$

The conditions of Assumption 4 are needed to prove the validity of an asymptotic approximation in Lemma 5 for the probability density of the discrete Fourier transform of the innovations ϵ_t. The first line is a Cramér condition. The second condition is used to approximate the probability density and it would not be necessary to approximate the probability distribution function. It implies that the probability distribution of ϵ_t has a bounded continuous density (see, for example, Theorem 3 in p. 509 of Feller (1971)).

3 Consistency

In this section we will consider the estimate $\widehat{d}$ when finite averages (for J fixed) of the periodogram of X_t are used under the linear process condition of Assumption 3. As we commented in the Introduction, the consistency proof is based on the approximation of the logarithm of the periodogram of X_t by that of ϵ_t, times the transfer function $|\alpha(\lambda)|^2$. This approximation will depend on the properties of the filter $\{\alpha_j\}$ and on the distribution of the linear innovations ϵ_t. Special care is needed because of the singularity of the logarithm function at the origin.

In the proofs we are able only to deal with the case $J \geq 2$. The reason of this limitation is the following. To approximate the periodogram of X_t by that of ϵ_t we need to consider the inverse moments of the periodogram of ϵ_t at the frequencies considered. The average of J periodogram ordinates of an i.i.d sequence will be asymptotically distributed as a χ^2_{2J} (up to constants). The key point is that if $Z \sim \chi^2_{2J}$, $E[Z^{-1}] < \infty$ for $J \geq 2$ (see Lemma 1 below). Of course, to approximate the moments of a random variable we need something more that its asymptotic distribution. That is why we approximate the probability density of the Fourier transform of ϵ_t and the regularity conditions on Assumption 4. We conjecture that a related argument can be used to construct a proof for $J = 1$.

We introduce some more notation. Write $I_j = I(\lambda_j)$ and $f_j = f(\lambda_j)$, and for the periodogram of the ϵ_t sequence, $I_{\epsilon j} = I_\epsilon(\lambda_j)$. Let J be a given, fixed, integer greater than or equal to 2. Define

$$\overline{I}_k = \sum_{j=1}^{J} I_{k+j-J} \qquad\qquad k = \ell + J, \ell + 2J, \ldots, m.$$

and equivalently

$$\overline{I}_{\epsilon k} = \sum_{j=1}^{J} I_{\epsilon, k+j-J} \qquad\qquad k = \ell + J, \ell + 2J, \ldots, m.$$

We suppress the dependence on J in the notation $\overline{I}_k$ and $\overline{I}_{\epsilon k}$.

Lemma 1 *Under Assumption 4 , for $J \geq 2$, $k \neq 0$ (mod N),*

$$E\left[\{\overline{I}_\epsilon(\lambda_k)\}^{-1}\right] < \infty.$$

We will make direct use of some results of Robinson (1995a and 1995b) to analyze the characteristics of the linear filter $\alpha(\lambda)$ under Assumptions 1 and 2. We can write, following Comte and Hardouin (1995a and 1995b),

$$\log \overline{I}_k = \log f_k + \log 2\pi \overline{I}_{\epsilon k} + \log(1 + F_k) + \log\left(1 + \frac{\delta_k}{H_k}\right), \tag{2}$$

where

$$\begin{aligned} F_k &= \frac{\sum_{j=1}^{J} I_{\epsilon,k+j-J}\left[f_{k+j-J} - f_k\right]}{\overline{I}_{\epsilon k}\, f_k} \\ H_k &= \sum_{j=1}^{J} 2\pi I_{\epsilon,k+j-J} f_{k+j-J} \\ \delta_k &= \overline{I}_k - H_k. \end{aligned}$$

We are interested in bound in probability the last two terms in equation (2), but first we prove a lemma that allows us to take logs of $\overline{I}_{\epsilon k}$ (and therefore of $\overline{I}_k$) and to divide by $\overline{I}_{\epsilon k}$.

Lemma 2 *Under Assumption 4, $J \geq 2$ and $k \neq 0$, $\overline{I}_{\epsilon k} > 0$, w.p.1.*

Then we have for ℓ increasing with N,

Lemma 3 *Under Assumptions 1, 2, 3 and 4,*

$$\log(1+F_k) = O_P(k^{-1}), \qquad\qquad k = \ell+J, \ell+2J, \ldots, m.$$

Lemma 4 *Under Assumptions 1, 2, 3 and 4, $J \geq 2$,*

$$\log\left(1+\frac{\delta_k}{H_k}\right) = O_P\left(\left[\frac{\log m}{k}\right]^{1/2}\right), \qquad\qquad k = \ell+J, \ell+2J, \ldots, m.$$

After this series of lemmas, we are in conditions of proving the consistency of $\widehat{d}$ for linear, possibly non-Gaussian, series under conditions 3 and 4. First, we introduce the following condition on the bandwidth numbers.

Assumption 5 *As $N \to \infty$,*

$$\frac{\log m}{\ell} + \frac{\ell\,(\log N)^2}{m} + \frac{m}{N} + \frac{\log m\,(\log N)^2}{m} \to 0.$$

This assumption is almost minimal given the structure of the estimate $\widehat{d}$. Then our first result is

Theorem 1 *Under Assumptions 1, 2, 3, 4 and 5, with $J \geq 2$, $\widehat{d} \to_P d$.*

The asymptotic distribution of the estimate is of evident interest, but the previous results are not enough for that. First, it is necessary to improve the approximation results between the periodogram ordinates of the observables and of the innovations. Then, a central limit theorem has to be proved for the random variable

$$\xi_N = \left(\sum_k \Lambda_k^2\right)^{-1}\left(\sum_k \Lambda_k \log 2\pi \overline{I}_{\epsilon k}\right),$$

that appears in the proof of Theorem 1. In the next section we propose tapering as a way of obtaining the above mentioned approximation and then we will investigate the asymptotic distribution of $\widehat{d}$.

4 Tapered discrete Fourier transform

In the previous section, we had obtained the consistency of the estimate with a pooling of periodogram ordinates. However, the bias of the periodogram makes impossible to obtain the asymptotic distribution from the proof of Theorem 1, unless we translate the regression to frequencies λ_j, $\ell+1 \leq j \leq \ell+m$, with $m/\ell \to 0$. Similar problem was observed for Gaussian series under stronger conditions by Comte and Hardouin (1995b, Propositions 1 and 3).

Tapering the data is a well known method to reduce the leakage in the periodogram from other frequencies, and in this case it is a very effective way of reducing the bias of the periodogram. We will need to strengthen Assumptions 1 and 2 on f to use these properties of tapering as in Assumption 3 of Robinson (1994b), with $1 < \alpha \leq 2$:

Assumption 6 *Further to Assumptions 1 and 2, denoting $g(\lambda) = G\lambda^{-2d}$, we assume, for some $0 < E_\alpha < \infty$, that as $\lambda \to 0^+$,*

$$\frac{f(\lambda)}{g(\lambda)} = 1 + E_\alpha \cdot \lambda^\alpha + o(\lambda^\alpha) \qquad\qquad 1 < \alpha \leq 2.$$

Assumption 6 is satisfied with $\alpha = 2$ by the fractional ARIMA and fractional noise models. This condition is equivalent to assume that $f(\lambda) = g(\lambda)h(\lambda)$, with $h(0) = 1$ and where h is differentiable with derivative in Lip(α-1) for $1 < \alpha \leq 2$.

Therefore this assumption will be satisfied at frequencies $\lambda_j = 2\pi j/N$, $j = 1, 2, \ldots, m$, for N big enough. Then, for frequencies $|\lambda| \leq \lambda_j/2$, we can expand f in this way:

$$f(\lambda_j - \lambda) = f(\lambda_j) - \lambda \cdot f'(\lambda_j) + O(\lambda_j^{-2d-\alpha}|\lambda|^\alpha), \tag{3}$$

where the derivative of f satisfies $f'(\lambda_j) = O(\lambda_j^{-2d-1})$. This can be seen heuristically in the following way: making a Taylor expansion of f around λ_j, we are led to study the difference for $|\theta| \leq 1$ and $\lambda \in [\lambda_j/2, 3\lambda_j/2]$,

$$|\lambda f'(\lambda_j - \theta\lambda) - \lambda f'(\lambda_j)|. \tag{4}$$

Now, as we can write $f' = h'g + g'h$, this is not bigger than the sum of the differences (taking $\theta = 1$ to simplify notation, w.l.o.g.)

$$\begin{aligned} & |h'(\lambda_j - \lambda)g(\lambda_j - \lambda) - h'(\lambda_j)g(\lambda_j)| && (5)\\ + \quad & |h(\lambda_j - \lambda)g'(\lambda_j - \lambda) - h(\lambda_j)g'(\lambda_j)|, && (6) \end{aligned}$$

times $|\lambda|$. First, (5) is bounded by

$$|h'(\lambda_j - \lambda) - h'(\lambda_j)|\,|g(\lambda_j - \lambda)| + |g(\lambda_j - \lambda) - g(\lambda_j)|\,|h'(\lambda_j)|,$$

and using the mean value theorem and that for these values of λ, $g = O(\lambda_j^{-2d})$, $g' = O(\lambda_j^{-2d-1})$, $g'' = O(\lambda_j^{-2d-2})$, $h = O(1)$, $h' = O(\lambda_j^{\alpha-1})$, and $|h'(\lambda_j - \lambda) - h'(\lambda_j)| = O(|\lambda|^{\alpha-1})$, this is bounded by

$$O\left(\lambda_j^{-2d}\lambda^{\alpha-1} + \lambda_j^{-2d-1}\lambda_j^{\alpha-1}\lambda\right) = O(\lambda_j^{-2d}\lambda^{\alpha-1}),$$

since $\alpha \in (1, 2]$. Similarly, (6) is bounded by

$$|g'(\lambda_j - \lambda) - g'(\lambda_j)|\,|h(\lambda_j - \lambda)| + |h(\lambda_j - \lambda) - h(\lambda_j)|\,|g'(\lambda_j)|,$$

and this is

$$O\left(\lambda_j^{-2d-2}\lambda + \lambda_j^{\alpha-1}\lambda_j^{-2d-1}\lambda\right) = O\left(\lambda_j^{-2d-\alpha}\lambda^{\alpha-1}\right).$$

Then (4) is easily seen now as $O(\lambda_j^{-2d-\alpha}\lambda^\alpha)$, multiplying the last two bounds by λ.

We will consider the full cosine bell or hanning taper, as suggested by Hurvich and Ray (1995) for a related problem. A generalization of the following results is ready available for any smooth data taper, but the hanning tapering has some desirable features that we will use later. Tapering allows us to reduce the bias of the periodogram for frequencies close to the origin if we assume a spectral density smooth enough at these frequencies (i.e., $\alpha > 1$ in Assumption 6). Also, since the tapered Fourier transform can be written down as a linear combination of three (plain) Fourier transforms, we can still use Chen and Hannan (1980) results as before, with minor modifications.

The tapered discrete Fourier transform is

$$w^T(\lambda_j) = \left(2\pi \sum h_t^2\right)^{-1/2} \sum_{t=1}^{N} h_t X_t \exp(i\lambda_j t),$$

where $h_t = \frac{1}{2}(1 - \cos[2\pi t/N])$, and the sum of the squared taper weights is $\sum h_t^2 = 3N/8$. This is called the *asymmetric* version of the cosine bell by Percival and Walden (1993, p. 325). The usual discrete Fourier transform $w(\lambda)$ is obtained setting $h_t \equiv 1$, $\forall t$.

Then, we can write (see Bloomfield (1976, pp. 80-84) or Percival and Walden (1993, pp. 325-326)), $2 \leq j \leq N-2$ the tapered Fourier transform at λ_j as a linear combination of the usual Fourier transform at the frequencies λ_j, λ_{j-1} and λ_{j+1},

$$w^T(\lambda_j) = \frac{1}{\sqrt{6}} \left[-w(\lambda_{j-1}) + 2w(\lambda_j) - w(\lambda_{j+1})\right].$$

The spectral kernel for the tapered periodogram, corresponding to Fejér kernel $K(\lambda)$ for raw the periodogram, is

$$\begin{aligned} K^T(\lambda_j - \lambda) &= \frac{1}{2\pi \sum h_t^2} \left| \sum_{t=1}^{N} h_t \exp\{it(\lambda_j - \lambda)\} \right|^2 = \frac{1}{2\pi \sum h_t^2} \left| D^T(\lambda_j - \lambda) \right|^2 \\ &= \frac{1}{2\pi \sum h_t^2} \sin^2[n(\lambda_j - \lambda)/2] H_j^2(\lambda), \end{aligned}$$

where

$$H_j(\lambda) = \frac{1}{\sqrt{6}} \left\{ \frac{2}{\sin[(\lambda_j - \lambda)/2]} - \frac{1}{\sin[(\lambda_{j-1} - \lambda)/2]} - \frac{1}{\sin[(\lambda_{j+1} - \lambda)/2]} \right\},$$

and $D^T(\lambda)$ is the equivalent of the Dirichlet kernel $D(\lambda)$ in the non tapered case. Obviously

$$D^T(\lambda_j) = \frac{1}{\sqrt{6}} \left\{2D(\lambda_j) - D(\lambda_{j-1}) - D(\lambda_{j+1})\right\}.$$

As a consequence, the tapered Fourier transform will not be uncorrelated, even asymptotically, for Fourier frequencies which are less than twice $\lambda_1 = 2\pi/N$ away, since otherwise in both transforms there is a common component. For that reason we will only consider frequencies λ_j and λ_k such that $k < j-2$, which we may expect to be uncorrelated, as in the general case. However, tapering will allow us to obtain much more neat results than previously, specially for the expectation of the periodogram, using expansion (3).

These improved characteristics derive from the following properties of K^T. It can be seen, that $K^T(\lambda)$ is even, positive, integrates to one and satisfies (see, e.g., Bloomfield (1976) or Hannan (1970, p. 265)):

- $\sup_{\lambda,N} |K^T(\lambda)| = O(N)$.
- $\sup_{\lambda,N} |K^T(\lambda)| = O(N^{-5}|\lambda|^{-6})$.

These properties follow from the fact that $\sup_{\lambda,N} |D^T(\lambda)| = O(\min\{N, N^{-2}|\lambda|^{-3}\})$. In words, the tails of K^T are less thicker than those of the non-tapered Fejér kernel K, but the central lobe is much wider. As both integrate to one, there is a trade off between the behaviour of the kernels at the origin and at the tails.

From the second property of K^T, the tapered periodogram will have improved asymptotic properties with respect to the usual periodogram, since the tails of the kernel $K^T(\lambda)$ decrease much faster with the frequency and with the sample size than the tails of Fejér kernel. Therefore, we will be able to reduce the bias of the periodogram on the tails, even for frequencies close to a singularity.

As we have seen, tapering destroys the orthogonality relations between Fourier transforms at different frequencies if they are too close. However, using the properties of the Dirichlet kernel $D(\lambda)$, have that, for $3 \leq j + k \leq N - 3$,

$$\int_{-\pi}^{\pi} D^T(\lambda_j - \lambda) D^T(\lambda + \lambda_k) d\lambda = 0. \tag{7}$$

In fact, this property holds for $D(\lambda)$ for $1 \leq j + k \leq N - 1$, but will not hold for any frequencies λ_j and λ_k and general taper weighting schemes.

We now present the equivalent of Theorem 2 of Robinson (1995a) for the (univariate) tapered Fourier transform. Define $v^T(\lambda) = w^T(\lambda)/(G^{1/2}\lambda^{-d})$.

Theorem 2 *Under Assumption 6 [$1 < \alpha \leq 2$], for any sequence of positive integers $j = j(N)$ and $k = k(N)$ such that $2 \leq k < j + 2 \leq N - 2$ and $j/N \to 0$ as $N \to \infty$,*

(a) $E[v^T(\lambda_j)\overline{v^T}(\lambda_j)] = 1 + O\left(j^{-\alpha} + [j/N]^{\alpha}\right)$,

(b) $E[v^T(\lambda_j)v^T(\lambda_j)] = O\left(j^{-4}\right)$,

(c) $E[v^T(\lambda_j)\overline{v^T}(\lambda_k)] = O\left(k^{-1}\right)$,

(d) $E[v^T(\lambda_j)v^T(\lambda_k)] = O\left(k^{-1}\right)$.

Comparing with Robinson's results for the expectation of the periodogram, in (a) we have improved the bound from $O(\log j/j)$ to $O(j^{-\alpha})$ for $1 < \alpha \leq 2$. This is the main bias reduction gain. The magnitude of this bound is determined by Assumption 6, and depends on the tapering, which makes all the other contributions of smaller order. This is the reason why we have such an improved bound in part (b). This improvement will be fundamental to approximate the (tapered and pooled) periodogram of the X_t sequence by the transfer function $|\alpha(\lambda)|$ times the periodogram of the innovations.

However, the bounds in Theorem 2 do not improve substantially for the correlations between Fourier transforms at different frequencies (just by a logarithm factor), since the frequencies can be arbitrarily close and tapering will not affect substantially the asymptotic behaviour of the periodogram there. Improved bounds are possible directly from the proof of the theorem if we consider the difference $|j-k|$.

5 Asymptotic distribution

In this section we derive the asymptotic distribution of $\widehat{d}$, for $J \geq 2$, when we use the tapered periodograms. First we need to modify the definition of the estimate in this way: define for $J = 1, 2, \ldots$, fixed, (assuming $(m-\ell)/(3J)$ integer),

$$Y_k^{(T,J)} = \log\left(\overline{I}^T(\lambda_k)\right) \qquad k = \ell + 3J,\ \ell + 6J,\ \ldots, m.$$

where

$$\overline{I}^T(\lambda_k) = \sum_{j=1}^{J} I^T(\lambda_{k+3(j-J)})$$

and

$$\widehat{d}^T = \left(\sum_k \Lambda_k^2\right)^{-1} \left(\sum_k \Lambda_k Y_k^{(T,J)}\right).$$

The reason for these definitions is immediate from the discussion in the previous sections. We consider a pooling or J tapered periodogram ordinates, and though each of the tapered Fourier transform is a linear combination of the Fourier transform at three adjacent frequencies, with this definition of $\widehat{d}^T$ we secure the asymptotic uncorrelatedness of $\overline{I}^T(\lambda_k)$ and asymptotic independence of the regressors $Y_k^{(T,J)}$ at different frequencies.

Let us introduce the following condition concerning the bandwidth numbers:

Assumption 7 *As $N \to \infty$,*

$$\frac{1}{\ell} + \frac{\ell(\log N)^2}{m} + \frac{m^{1+1/2\alpha}}{N} + \frac{\log N \log m}{m^{(\alpha-1)/2}} \to 0.$$

The basic different with respect to Assumption 6 of Robinson (1995a) is the first condition (just $\ell \to \infty$): we only need the trimming number ℓ to increase as slow as we want, since from our Theorem 2 we can control the bias of the periodogram for closer frequencies to the origin thanks to tapering. Then, we present our main result:

Theorem 3 *Under Assumptions 3, 4, 6, and 7, if ϵ_t has moments of all orders, $J \geq 2$,*

$$m^{1/2}\left(\widehat{d}^T - d\right) \to_d N(0, \frac{3J}{4}\psi'(J)),$$

where $\psi'(x) = \frac{d}{dx}\psi(x) = \frac{d}{dx}\log\Gamma(x)$ is the digamma function.

The proof of the theorem is based on the method of moments. Although for the estimation of the moments of the logarithm of each of the innovations (pooled and tapered) periodogram ordinates we only require four bounded moments of ϵ_t (three could suffice for some applications), this is not enough to approximate the moments of a normalized infinite average of such periodogram logarithms. Our moment assumption is then used to approximate with enough degree of accuracy those moments by means of Edgeworth expansions for the probability density of the Fourier transform.

For the asymptotic normality proof we do not use any special properties of tapering or pooling the periodogram, apart from an approximation equivalent to Lemma 4 (pooling) and Theorem 2 (tapering). These two devices are used to improve the approximations and behaviour of the periodogram of the long range dependent time series. Possibly, under stronger conditions on the dependence of the process X_t and/or its distribution one or both of these techniques could be dispensed with.

6 Simulation work

In this section we present a limited simulation exercise to analyze the techniques of tapering and pooling in the log-periodogram estimate for non-Gaussian data. To that end we simulate 5000 series following an ARFIMA$(0,d,0)$ model, and innovations with a t_5 distribution, which only has four moments, so Theorem 1 holds, but not Theorem 3. We report only the results for $d = .45$, the conclusions being similar for other values in the interval $(0, \frac{1}{2})$. We use for the simulations a modification of the function `arima.fracdiff.sim` included in the package SPLUS.

The sample size is $N = 512$ and the bandwidth numbers considered are $m = 30, 60, 90$, and the pooling numbers $J = 1, 2, 3$. We do not perform any trimming, $\ell = 0$, this not being a decisive choice. For each of the time series simulated we calculate three different types of estimate for the all nine combinations of bandwidth and pooling choices. They are the non-tapered log-periodogram $\widehat{d}$, the tapered log-periodogram $\widehat{d}^T$, as defined previously, and a modification of this last one, $\widehat{d}^{T,\star}$, considering all possible frequencies between the origin and λ_m, with

$$\overline{I}^T(\lambda_k) = \sum_{j=1}^{J} I^T(\lambda_{k+j-J}), \qquad k = \ell + J, \ell + 2J, \ldots, m.$$

In Figures 1 to 3 we give the boxplots of the replications and we report in Tables I to III bellow the results of the simulations for the three estimates. We give for all the estimates calculated the bias across replications, the standard deviation, the asymptotic standard deviation in the appropriate central limit theorem (CLT), the mean square error and the true coverage in the 90%, 95% and 99% confidence intervals calculated using the previous CLT's. For the third estimate we do not provide asymptotic theory, although its consistency can be shown using the same techniques. The main difficulty here is that, without the additional spacing between regressors, we can not guarantee their asymptotic independence and the approach used in our proofs breaks down.

Following the discussion in Robinson (1995a), letting J increase may produce asymptotic efficiency gains, since $J\psi'(J)$ is decreasing. This can be checked in the column for the *theoretical* standard deviation, *th.sd.* However, in practice and for this sample size, the gains are only apparent for the log-periodogram without tapering and $J = 2$ but not for $J = 3$, where due to the reduced number of regressors, the estimates have now bigger variances. When we taper the observations, $\widehat{d}^T$ is already using a smaller number of frequencies with $J = 1$, so to use $J > 1$ always increases the variance. The situation is much different when we do not space the regressors, obtaining efficiency gains with larger values of J, and with much reduced variances than $\widehat{d}^T$, but bigger than those we would expect if the regressors were uncorrelated (or independent), reported in the column *th.sd** in Table III.

In all cases considered, the variance decreases with m, and also the bias, since for these series the semiparametric model considered is a good approximation for all the range of frequencies $(0, \pi)$. The bias, in general, tends to increase with J. This leads to a minimum mean square error (MSE) for the estimates with biggest m. There are not big differences in MSE between tapered and not tapered when all the possible frequencies are considered and only a slight increment when we consider asymptotically independent regressors in $\widehat{d}^T$.

The accuracy of the CLT deteriorates with tapering, since we are effectively reducing the number of observations in the log-periodogram regression. The same argument may apply to the estimates with higher J, for which the confidence intervals are quite imprecise, specially for tapered estimates. In general, it seems that with such a flat distribution, very prone to the presence of outliers, the asymptotics would need larger samples sizes to provide equivalently precise approximations as, for example, with Gaussian data.

Table I. Log-Periodogram estimate. No taper. $N = 512$, $\epsilon_t \sim t_5$. 5000 Replications

m	J	d	bias	sd	th.sd	MSE	90% Cov.	95% Cov.	99% Cov.
30	1	.45	0.0113	0.1415	0.1170	0.2329	82.82	89.20	96.66
	2	.45	-0.0244	0.1174	0.1036	0.1948	85.12	91.56	97.18
	3	.45	0.0335	0.1298	0.0993	0.2506	77.28	85.04	93.94
60	1	.45	.0048	0.0930	0.0820	0.2156	85.72	91.46	97.66
	2	.45	-0.0199	0.0793	0.0732	0.1912	86.64	92.32	97.88
	3	.45	0.0179	0.0837	0.0702	0.2260	81.78	89.34	96.92
90	1	.45	-0.0019	0.0740	0.0675	0.2062	86.88	92.62	98.04
	2	.45	-0.0213	0.0631	0.0598	0.1877	86.28	92.30	97.78
	3	.45	0.0077	0.0655	0.0573	0.2138	84.70	90.88	97.54

Table II. Log-Periodogram estimate. Cosine taper. $N = 512$, $\epsilon_t \sim t_5$. 5000 Replications									
m	J	d	bias	sd	th.sd	MSE	90% Cov.	95% Cov.	99% Cov.
30	1	.45	0.0125	0.2520	0.2027	0.2774	82.86	89.00	96.20
	2	.45	-0.0002	0.2690	0.1795	0.2746	74.30	81.86	91.24
	3	.45	0.0641	0.3100	0.1721	0.3604	62.30	70.80	83.64
60	1	.45	0.0033	0.1638	0.1433	0.2323	85.96	91.78	97.28
	2	.45	-0.0156	0.1799	0.1269	0.2210	76.72	83.80	92.54
	3	.45	0.0241	0.1960	0.1217	0.2632	69.00	77.72	89.24
90	1	.45	-0.0011	0.1265	0.1170	0.2175	87.36	92.74	98.08
	2	.45	-0.0205	0.1437	0.1036	0.2050	77.12	84.28	93.42
	3	.45	0.0063	0.1404	0.0993	0.2279	76.02	83.92	92.78

Table III. Log-Periodogram estimate. Cosine taper*. $N = 512$, $\epsilon_t \sim t_5$. 5000 Replications						
m	J	d	bias	sd	th.sd*	MSE
30	1	.45	0.0097	0.2004	0.1170	0.2515
	2	.45	-0.0141	0.1714	0.1036	0.2193
	3	.45	0.0243	0.1847	0.0993	0.2591
60	1	.45	0.0006	0.1257	0.0827	0.2189
	2	.45	-0.0153	0.1100	0.0732	0.2010
	3	.45	0.0086	0.1139	0.0702	0.2233
90	1	.45	-0.0049	0.0969	0.0675	0.2074
	2	.45	-0.0180	0.0855	0.0598	0.1938
	3	.45	0.0002	0.0872	0.0573	0.2103

7 Appendix: Proofs of Section 3

The following is a simplified version of Chen and Hannan's Lemma 2, where we only use the first two terms of an Edgeworth expansion for the probability density of the Fourier transform of ϵ_t, so only four bounded moments are required.

Lemma 5 (Chen and Hannan (1980)) *Under Assumption 4, the probability distribution function Q_N of the vector*

$$W_N = N^{-1/2} \sum_{t=1}^{N} Y_t,$$

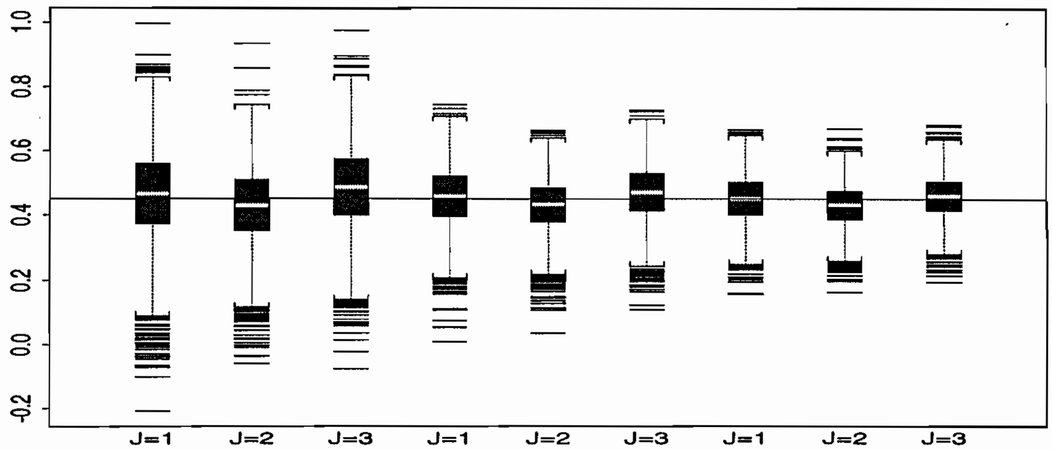

Figure 1: Log-periodogram estimate, t_5 ARFIMA$(0, .45, 0)$, $N = 512$, $m = 30, 60, 90$. 5000 Replications.

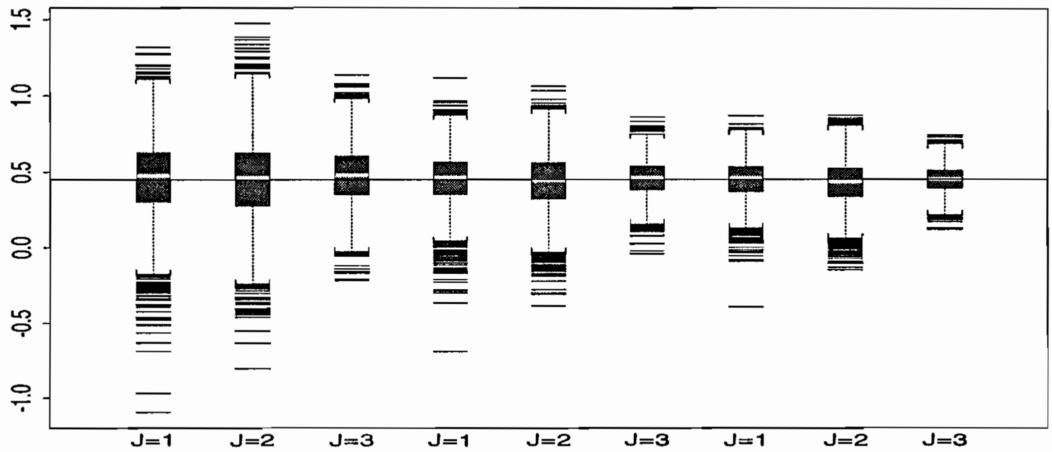

Figure 2: Tapered Log-periodogram estimate, t_5 ARFIMA$(0, .45, 0)$, $N = 512$, $m = 30, 60, 90$. 5000 Replications.

where

$$Y_t' = Y'(j(1), \ldots, j(k)) = \sqrt{2}\epsilon_t(\cos t\lambda_{j(1)}, \sin t\lambda_{j(1)}, \ldots, \cos t\lambda_{j(k)}, \sin t\lambda_{j(k)}),$$

has density q_N for all sufficiently large N and

$$\sup_{\mathbf{y}\in\mathcal{R}^k} (1 + \|\mathbf{y}\|^4) \left| q_N(\mathbf{y}) - \sum_{r=0}^{1} N^{-r/2} P_r(-\phi : \overline{\chi}_{\nu,N})(\mathbf{y}) \right| = O(N^{-1}), \tag{8}$$

where P_r are polynomials in the average of the joint cumulants of Y_t $(1 \le t \le N)$ of order $\nu = (\nu_1, \ldots, \nu_{2k})$, $\overline{\chi}_{\nu,N}$, multiplied by the $2k$th multivariate Normal density ϕ and where $P_0(\mathbf{y}) = \phi(\mathbf{y})$.

The previous result is used to prove Lemma 1 about the inverse moments of the periodogram of an i.i.d. sequence:

Proof of Lemma 1. First, from Lemma 5, $2\pi\overline{I}_{\epsilon k}$ has the probability density of a $\frac{1}{2}\chi^2_{2J}$ distribution with error (using only P_0) of order $O((1 + \|\mathbf{y}\|^4)^{-1}N^{-1/2})$. Also the density of a χ^2_{2J} is

$$\phi_{\chi^2_{2J}}(x) = \frac{x^{J-1}\, e^{-x/2}}{(J-1)!\, 2^J} \qquad 0 \le x < \infty.$$

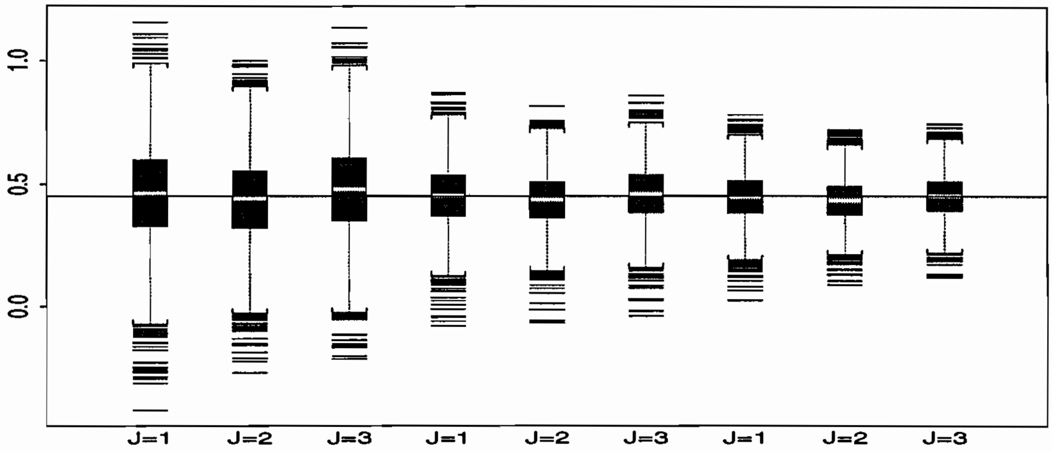

Figure 3: Tapered, no-spacing, Log-periodogram estimate, t_5 ARFIMA(0,.45,0), $N = 512$, $m = 30, 60, 90$. 5000 Replications.

It is clear that if $X \sim \chi^2_{2J}$ then $E[X^{-1}] < \infty$ if $J \geq 2$. Thus we only need to check that the error in the evaluation of the second inverse moment of $\overline{I}_{\epsilon k}$ using the Lemma 5 is bounded. If we write

$$\overline{I}_{\epsilon k} = \sum_{j=1}^{J} I_\epsilon(\lambda_{k+j-J}) = \sum_{j=1}^{J}(y_{aj}^2 + y_{bj}^2)$$

we need

$$N^{-1/2}\int_{\mathcal{R}^{2J}}(1+||\mathbf{y}||^4)^{-1}\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}d\mathbf{y} \; < \; \infty. \tag{9}$$

First, defining the sets $A = [-1,1]^{2J}$ and A^c its complementary in $\mathcal{R}^{2J}$,

$$\begin{aligned}\int_{\mathcal{R}^{2J}}(1+||\mathbf{y}||^4)^{-1}\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}d\mathbf{y} \;&\leq\; \text{const.}\int_A\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}d\mathbf{y}\\ &\quad+\text{const.}\int_{A^c}(1+||\mathbf{y}||^4)^{-1}d\mathbf{y}\\ &\leq\; \text{const.}\int_A\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}d\mathbf{y}+\text{const.},\end{aligned}$$

since $(1+||\mathbf{y}||^4)^{-1}$ and $\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}$ are bounded from above in A and A^c, respectively. Next, to bound the remaining integral, if $\phi(.)$ denote the densities of the correspondent distributions, we have

$$\begin{aligned}\infty > \int_0^\infty x^{-1}\phi_{\chi^2_{2J}}(x)dx \;&=\; \int_{\mathcal{R}^{2J}}\left(\sum_j(y_{aj}^2+y_{bj}^2)\right)^{-1}\phi_{\mathcal{N}(0,I_{2J})}(\mathbf{y})d\mathbf{y}\\ &>\; \int_A\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}\phi_{\mathcal{N}(0,I_{2J})}(\mathbf{y})d\mathbf{y}\\ &\geq\; \text{const.}\int_A\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}d\mathbf{y},\end{aligned}$$

as the Normal density is bounded from below in A. Hence, we have got

$$\int_A\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}d\mathbf{y} \leq \text{const.} < \infty.$$

Then, the l.h.s. of (9) is $O(N^{-1/2})$, and the Lemma follows.

An alternate way of checking that the error is actually $O(N^{-1/2})$ is bounding directly the integrals

$$\int_{\mathcal{R}^{2J}}(1+||\mathbf{y}||^4)^{-1}\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}d\mathbf{y} \leq \text{const.}\int_A\left(\sum(y_{aj}^2+y_{bj}^2)\right)^{-1}d\mathbf{y} \; + \; \text{const.}.$$

This can be done as follows. In parenthesis appear the code number of the integral used from Gradshteyn and Ryzhik (1980). In each step we denote as p the sum of the squares of the remaining variables with respect to we are not integrating. Then, using (3.241.4),

$$\int_0^\infty (x^2+p)^{-1}dx = \frac{\pi}{2}p^{-1/2},$$

and from (2.271.5)

$$\int_0^1 (x^2+p)^{-1/2}dx = \log(1+\sqrt{p+1}) - \frac{1}{2}\log p.$$

Now we have

$$\int_0^1 \left[\log(1+\sqrt{x+1}) - \frac{1}{2}\log x\right] dx < \infty$$

and the global integral is bounded; to make this process we have needed 3 integrals: that is the reason why we need $J \geq 2$ to get at least 4 degrees of freedom.

Proof of Lemma 2. Using Lemma 1 we have that, for any $\theta > 0$,

$$\begin{aligned} P\{\overline{I}_{\epsilon j} = 0\} &\leq P\{\overline{I}_{\epsilon k} \leq N^{-\theta}\} \\ &\leq P\{\overline{I}_{\epsilon j}^{-1} \geq N^{\theta}\} \\ &= O(N^{-\theta}), \end{aligned}$$

and the lemma follows from the Borel-Cantelli Lemma, choosing $\theta > 1$.

Proof of Lemma 3. For $j = 1+k-J, \ldots, k$ we have

$$\begin{aligned} \max_j |f_j - f_k| &\leq \max_j \sup_{\lambda\in[\lambda_k,\lambda_j]} |f'(\lambda)|\,|\lambda_j - \lambda_k| \\ &= O(f_k\,\lambda_k^{-1}\,N^{-1}) \\ &= O(f_k\,k^{-1}). \end{aligned}$$

Then we have, since $I_{\epsilon j} \geq 0$,

$$|F_k| \leq \frac{\max_j |f_j - f_k|\,\overline{I}_{\epsilon k}}{f_k \overline{I}_{\epsilon k}} = O(k^{-1}).$$

Then $F_k = \dot{O}_P(\ell^{-1})$, uniformly in k, and using $|\log(1+x)| \leq 2|x|$ for $|x| \leq 1/2$, we obtain as $\ell \to \infty$

$$|\log(1+F_k)| \leq 2|F_k| = O_P(k^{-1}).$$

Proof of Lemma 4. First, for summations running from $j = 1+k-J$ to $j = k$,

$$\begin{aligned} E[(H_k)^{-1}] &= E\left[\left(2\pi\sum\nolimits_j f_j I_{\epsilon j}\right)^{-1}\right] \\ &\leq (2\pi)^{-1}\left\{\max_j f_j^{-1}\right\} E\left[(\overline{I}_{\epsilon k})^{-1}\right] \\ &= O(f_k^{-1}), \end{aligned}$$

using Lemma 1. Now, from Robinson (1995b, expression (3.17)),

$$\begin{aligned}
E[|\delta_k|] &= E\left[\left|\sum_j I_j - 2\pi I_{\epsilon j} f_j\right|\right] \\
&\leq \sum_j E\left|I_j - 2\pi I_{\epsilon j} f_j\right| \\
&= O\left(\max_j f_j \left[\frac{\log j}{j}\right]^{1/2}\right) \\
&= O\left(f_k \left[\frac{\log k}{k}\right]^{1/2}\right).
\end{aligned}$$

Then the result follows from $\delta_k = O_P(f_k \left[\frac{\log k}{k}\right]^{1/2})$ and $(H_k)^{-1} = O_P(f_k^{-1})$, and the same reasoning of the previous lemma, since $|\delta_k/H_k| = o_P(1)$, uniformly in k.

Proof of Theorem 1. From Robinson (1995a) and the definition for the summation in k, we can obtain

$$\begin{aligned}
&\sum_k \Lambda_k^2 = \frac{4m}{J}\left(1 + O\left(\frac{\ell(\log N)^2}{m}\right)\right) = \frac{4m}{J}(1 + o(1)), \\
&\sup_{\ell \leq k \leq m} |\Lambda_k| = O(\log N) \\
&\sum_k |\Lambda_k|^p = O(m), \qquad p \geq 1.
\end{aligned} \tag{10}$$

Hence, under Assumption 1, with the previous properties,

$$\left(\sum_k \Lambda_k^2\right)^{-1} \sum_k \Lambda_k \log f_k = d + O\left(\left[\frac{m}{N}\right]^{\alpha}\right). \tag{11}$$

Now, from Lemmas 3 and 4 we have that

$$\log \overline{I}_k = \log f_k + \log \overline{I}_{\epsilon k} + O_P\left(\left[\frac{\log m}{k}\right]^{1/2}\right). \tag{12}$$

Substituting in the definition of $\widehat{d}$ and using (11), $\sum_k \Lambda_k^2 \sim 4m/J$ and $\sup|\Lambda_k| = O(\log N)$,

$$\begin{aligned}
\widehat{d} &= \left(\sum_k \Lambda_k^2\right)^{-1}\left(\sum_k \Lambda_k \log \overline{I}_k\right) \\
&= \left(\sum_k \Lambda_k^2\right)^{-1}\left(\sum_k \Lambda_k \left[\log f_k + \log 2\pi \overline{I}_{\epsilon k} + O_P\left(\left[k^{-1}\log m\right]^{1/2}\right)\right]\right) \\
&= d + \left(\sum_k \Lambda_k^2\right)^{-1}\left(\sum_k \Lambda_k \log 2\pi \overline{I}_{\epsilon k}\right) + O_P\left(\frac{\log N}{m}(\log m)^{1/2}\sum_k k^{-1/2}\right) \\
&\quad + O([mN^{-1}]^{\alpha}\log N) \\
&= d + \xi_N + O_P\left(\log N (\log m)^{1/2} m^{-1/2} + [mN^{-1}]^{\alpha}\right), \quad \text{say,} \\
&= d + \xi_N + o_P(1),
\end{aligned}$$

where the last line follows from Assumption 5.

To prove the consistency of the estimate $\widehat{d}$ we only need to calculate the first two moments of the variable

$$\xi_N = \left(\sum_k \Lambda_k^2\right)^{-1} \left(\sum_k \Lambda_k \log 2\pi \overline{I}_{\epsilon k}\right).$$

To evaluate the moments of $\overline{I}_{\epsilon k}$, we approximate the probability density of the Fourier transform, $q_N(\mathbf{y})$, using Chen and Hannan's Lemma 5. This result uses some results in Bhattacharya and Rao (1975) to approximate the density of the Fourier transform $w_\epsilon(\lambda)$ of the sequence ϵ_t. They employed a finite fifth moment of ϵ_t to get a stronger result. For our purposes with Lemma 5 is enough.

Set $4\pi\overline{I}_{\epsilon k} = \sum_{j=1}^{J}(y_{aj}^2 + y_{bj}^2)$, where y_{aj} and y_{bj} correspond to the sine and cosine summations, respectively, of $4\pi I_{\epsilon j}$. Now, from Chen and Hannan (1980), (see Lemma 5),

$$P_1(-\phi : \overline{\chi}_{\nu,N})(\mathbf{y}) = \sum_{|\nu|=3} \frac{\overline{\chi}_{\nu,N}}{v!} \prod_{j=1}^{2k} \left(\frac{\partial}{\partial y_i}\right)^{\nu_j} \phi(\mathbf{y}), \quad v! = \prod_j v_j!.$$

As $|\nu| = 3$ the terms in P_1 are of one of the following types when we are considering the joint distribution in $\mathcal{R}^{4J}$ of the sine and cosine components $4\pi\overline{I}_{\epsilon k}$ and $4\pi\overline{I}_{\epsilon k'}$, $k \neq k'$ (up to constants):

1. $H_3(y_s)\phi(\mathbf{y})$, where H_i are the Hermite polynomials of order i and $s \in \{1,\dots,4J\}$,

$$H_i(x)\phi(x) = (-1)^i \left(\frac{\partial}{\partial x}\right)^i \phi(x), \qquad x \in \mathcal{R}.$$

   Then this term is odd in the component y_s of $\mathbf{y}$ (since H_3 is odd and ϕ is even).

2. $H_2(y_s)H_1(y_r)\phi(\mathbf{y})$, $y_r \neq y_s$ and $r, s \in \{1,\dots,4J\}$. Then this term is odd in the component y_r.

3. $H_1(y_s)H_1(y_r)H_1(y_u)\phi(\mathbf{y})$, y_r, y_s, y_u all different. Then this term is odd in the components y_s, y_r and y_u.

If $k = k'$, we consider only a distribution in $\mathcal{R}^{2J}$ and the typical terms of P_1 are:

1. $H_3(y_s)\phi(\mathbf{y})$, where $s \in \{1,\dots,2J\}$. Then this term is odd in the component y_s of $\mathbf{y}$.

2. $H_2(y_s)H_1(y_r)\phi(\mathbf{y})$, $r \neq s$ and $r, s \in \{1,\dots,2\}$. Then this term is odd in the component y_r.

Then we have

$$\begin{aligned} E\left[\log 2\pi\overline{I}_{\epsilon k}\right] + \log 2 &= \int_{\mathcal{R}^{2J}} \log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) q_N(\mathbf{y}) d\mathbf{y} \\ &= \int_{\mathcal{R}^{2J}} \log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) \left[\phi(\mathbf{y}) + N^{-1/2}P_1(\mathbf{y})\right] d\mathbf{y} + O(N^{-1}) \\ &= \psi(J) + \log 2 + O(\frac{1}{N}), \end{aligned}$$

since $\int_0^\infty (\log x)^h/(1+x^5)dx < \infty$ and $\int_0^\infty (x\log x)^h e^{-x} dx < \infty$, for all $h \geq 0$. $\psi(z) = d/dz \log\Gamma[z]$ is the digamma function. The contribution from $P_1(\mathbf{y})$ is 0 since the interval of integration is $(-\infty, \infty)$ and P_1 is always odd in one component of $\mathbf{y}$ and the log term is even in all the components.

Consider now the Covariance terms. Denote $E_k = E[\log 2\pi \overline{I}_{\epsilon k}]$. Then $(k \neq k')$

$$
\begin{aligned}
&\operatorname{Cov}\left[\log 2\pi \overline{I}_{\epsilon k}, \log 2\pi \overline{I}_{\epsilon k'}\right] \\
&= \int_{\mathcal{R}^{4J}} \left[\log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]\left[\log\left(\sum\nolimits_{j'} (y_{aj'}^2 + y_{bj'}^2)\right) - E_{k'}\right] q_N(\mathbf{y})d\mathbf{y} \\
&= \int_{\mathcal{R}^{4J}} \left[\log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]\left[\log\left(\sum\nolimits_{j'} (y_{aj'}^2 + y_{bj'}^2)\right) - E_{k'}\right] \\
&\qquad \times \left[\phi(\mathbf{y}) + N^{-1/2}P_1(\mathbf{y})\right] d\mathbf{y} + O(N^{-1}) \\
&= N^{-1/2}\int_{\mathcal{R}^{4J}}\left[\log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]\left[\log\left(\sum\nolimits_{j'} (y_{aj'}^2 + y_{bj'}^2)\right) - E_{k'}\right]P_1(\mathbf{y}) + O(\frac{1}{N}) \\
&= O(N^{-1}),
\end{aligned}
$$

as $\phi(\mathbf{y})$ is the density of the standard Normal density in $\mathcal{R}^{4J}$ (with uncorrelated components!), and since the contribution from P_1 cancel out by the same argument as before.

Now for the variance we have:

$$
\begin{aligned}
\operatorname{Var}\left[\log 2\pi \overline{I}_{\epsilon k}\right] &= \int_{\mathcal{R}^{2J}} \left[\log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]^2 q_N(\mathbf{y})d\mathbf{y} \\
&= \int_{\mathcal{R}^{2J}}\left[\log\left(\sum\nolimits_j (y_{aj}^2 + y_{bj}^2)\right) - E_k\right]^2\left[\phi(\mathbf{y}) + N^{-1/2}P_1(\mathbf{y})\right]d\mathbf{y} + O(\frac{1}{N}) \\
&= \psi'(J) + O(N^{-1}),
\end{aligned}
$$

reasoning as before.

Then it is immediate that, using (10),

$$E[\xi_N] = O(N^{-1}),$$

and that,

$$\operatorname{Var}[\xi_N] = \frac{J}{4m}\psi'(J) + O(N^{-1}) + o(m^{-1}) \sim \frac{J}{4m}\psi'(J).$$

Therefore $\xi_N = o_P(1)$ with Assumption 5 and the theorem is proved.

8 Appendix: Proof of Theorem 2

First, we study the bias of the tapered periodogram $\mathrm{I}^T(\lambda_j) = |v^T(\lambda_j)|^2$ with respect to $f(\lambda_j)$. We use exactly the same method of proof as Robinson (1995a), using the improved properties of $K^T(\lambda)$ with respect to Fejér kernel. The additional term $O([j/N]^\alpha)$ that shows up when we normalize with respect to $G\lambda_j^{-2d}$ (instead that with respect to $f(\lambda_j)$) follows as in that paper.

Given that the expectation of the tapered periodogram is now

$$E[|w^T(\lambda_j)|^2] = \int_{-\pi}^{\pi} f(\lambda)K^T(\lambda_j - \lambda)d\lambda,$$

we consider the same intervals of integration to analyze the bias in

$$E[|w^T(\lambda_j)|^2] - f(\lambda_j) = \int_{-\pi}^{\pi} [f(\lambda) - f(\lambda_j)]\, K^T(\lambda_j - \lambda)d\lambda,$$

as Robinson (1995a). Consider a fixed $\epsilon > 0$, such that $f(\lambda) \leq C_\epsilon \lambda^{-2d}$, $|\lambda| \in (0, \epsilon)$ for some positive constant C_ϵ, depending on ϵ, and N big enough such that $\lambda_j, \lambda_k < \epsilon$. Then,

$$\left| \int_{-\pi}^{-\epsilon} + \int_{\epsilon}^{\pi} \right| \leq \max_{|\lambda| \geq \epsilon} \left| K^T(\lambda_j - \lambda) \right| \int_{-\pi}^{\pi} |f(\lambda) - f(\lambda_j)| \, d\lambda = O(f(\lambda_j) \cdot N^{-5}),$$

using the properties of $K^T(\lambda)$ and the integrability of f, by covariance stationarity. Next,

$$\begin{aligned} \left| \int_{-\epsilon}^{-\lambda_j/2} \right| &\leq \left[\max_{\lambda_j/2 \leq \lambda \leq \epsilon} f(\lambda) + f(\lambda_j) \right] \int_{-\pi}^{-\lambda j/2} \left| K^T(\lambda_j - \lambda) \right| d\lambda \\ &= O\left(f(\lambda_j) \cdot N^{-5} \int_{\lambda_j/2}^{\pi} \lambda^{-6} d\lambda \right) = O\left(f(\lambda_j) \cdot j^{-5} \right). \end{aligned}$$

Identical bound can be obtained for the interval $[3\lambda_j/2, \epsilon]$. Now

$$\begin{aligned} \left| \int_{-\lambda_j/2}^{\lambda_j/2} \right| &\leq \left[\max_{-\epsilon \leq \lambda \leq -\lambda_j/2} \left| K^T(\lambda_j - \lambda) \right| \right] \int_{-\lambda_j/2}^{\lambda_j/2} (f(\lambda) + f(\lambda_j)) \, d\lambda \\ &= O\left(N^{-5} \lambda_j^{-6} \left(\int_0^{\lambda_j} \lambda^{-2d} d\lambda + \lambda_j^{1-2d} \right) \right) = O\left(f(\lambda_j) \cdot j^{-5} \right). \end{aligned}$$

Next, using (3),

$$\begin{aligned} \left| \int_{\lambda_j/2}^{3\lambda_j/2} [f(\lambda) - f(\lambda_j)] \, K^T(\lambda_j - \lambda) d\lambda \right| &= \left| \int_{-\lambda_j/2}^{\lambda_j/2} [f(\lambda_j - \lambda) - f(\lambda_j)] \, K^T(\lambda) d\lambda \right| \\ &= \left| \int_{-\lambda_j/2}^{\lambda_j/2} \left[\lambda \cdot f'(\lambda_j) + O(\lambda_j^{-\alpha - 2d} |\lambda|^\alpha) \right] K^T(\lambda) d\lambda \right| \\ &= O\left(\lambda_j^{-\alpha - 2d} \int_{-\lambda_j/2}^{\lambda_j/2} |\lambda|^\alpha K^T(\lambda) d\lambda \right), \end{aligned}$$

since K^T is even and we are integrating in a symmetric interval around 0. Now, with $\alpha \in (1, 2]$,

$$\begin{aligned} \int_{-\lambda_j/2}^{\lambda_j/2} |\lambda|^\alpha K^T(\lambda) &= 2 \left\{ \int_0^{N^{-1}} + \int_{N^{-1}}^{\lambda_j/2} \right\} \lambda^\alpha K^T(\lambda) d\lambda \\ &= 2 \left\{ \int_0^{N^{-1}} + \int_{N^{-1}}^{\lambda_j/2} \right\} \lambda^\alpha K^T(\lambda) d\lambda \\ &= O\left(N \int_0^{N^{-1}} \lambda^\alpha d\lambda + N^{-5} \int_{N^{-1}}^{\lambda_j/2} \lambda^{\alpha - 6} d\lambda \right) \\ &= O(N^{-\alpha}). \end{aligned}$$

Therefore

$$\left| \int_{\lambda_j/2}^{3\lambda_j/2} \right| = O\left(\lambda_j^{-\alpha - 2d} \cdot N^{-\alpha} \right) = O(f(\lambda_j) \cdot j^{-\alpha}).$$

Let's study the covariances between tapered Fourier transforms of X_t. As before, we are led to the expression, $j = 1$

$$E[w^T(\lambda_j)^2] = \frac{1}{2\pi \sum h_t^2} \int_{-\pi}^{\pi} D^T(\lambda_j - \lambda) D^T(\lambda_j + \lambda) f(\lambda) d\lambda$$

which is equal to

$$\frac{1}{2\pi \sum h_t^2} \int_{-\pi}^{\pi} [f(\lambda) - f(\lambda_j)] \, D^T(\lambda_j - \lambda) D^T(\lambda_j + \lambda) f(\lambda) d\lambda, \tag{13}$$

using (7). Now, we can study the integral in (13) splitting the range of integration in the following intervals,

$$\left|\int_{-\pi}^{-\epsilon}+\int_{\epsilon}^{\pi}\right| = O\left(\frac{1}{N^5\epsilon^6}\int_{-\pi}^{\pi}[f(\lambda)+f(\lambda_j)]\,d\lambda\right) = O(f(\lambda_j)\cdot N^{-5}),$$

$$\begin{aligned}\left|\int_{-\epsilon}^{-2\lambda_j}+\int_{2\lambda_j}^{\epsilon}\right| &= O\left(\left[\max_{2\lambda_j\leq\lambda\leq\epsilon} f(\lambda)+f(\lambda_j)\right]N^{-1}\int_{2\lambda_j}^{\pi}\left|D^T(\lambda_j-\lambda)D^T(\lambda_j+\lambda)\right|d\lambda\right)\\ &= O\left(f(\lambda_j)\cdot N^{-5}\int_{2\lambda_j}^{\pi}\lambda^{-6}d\lambda\right) = O\left(f(\lambda_j)\cdot j^{-5}\right).\end{aligned}$$

Now, using $f(\lambda_j)=f(-\lambda_j)$

$$\begin{aligned}\left|\int_{-2\lambda_j}^{-\lambda_j/2}+\int_{\lambda_j/2}^{2\lambda_j}\right| &= O\left(N^{-1}\sup_{\lambda_j/2\leq\lambda\leq2\lambda_j}|f'(\lambda)D^T(\lambda_j+\lambda)|\int_{\lambda_j}^{2\lambda_j}|\lambda_j-\lambda||D^T(\lambda_j-\lambda)|d\lambda\right)\\ &= O\left(N^{-1}\cdot f(\lambda_j)\lambda_j^{-1}\cdot N^{-2}\lambda_j^{-3}\int_0^{2\lambda_j}\lambda|D^T(\lambda)|d\lambda\right) = O(f(\lambda_j)\cdot j^{-4}),\end{aligned}$$

because

$$\int_0^{2\lambda_j}\lambda|D^T(\lambda)|d\lambda = O\left(N\int_0^{N^{-1}}\lambda d\lambda + N^{-2}\int_{N^{-1}}^{2\lambda_j}\lambda^{-2}d\lambda\right) = O(N^{-1}),$$

using the properties of $D^T(\lambda)$. Finally

$$\begin{aligned}\left|\int_{-\lambda_j/2}^{\lambda_j/2}\right| &= O\left(\max_{-\lambda_j/2\leq\lambda\leq\lambda_j/2}\left|D^T(\lambda_j-\lambda)D^T(\lambda_j+\lambda)\right|\int_{-\lambda_j/2}^{\lambda_j/2}[f(\lambda)+f(\lambda_j)]\,d\lambda\right)\\ &= O\left(N^{-5}\lambda_j^{-6}\left[\int_0^{\lambda_j/2}\lambda^{-2d}d\lambda+\lambda_j f(\lambda_j)\right]\right) = O(f(\lambda_j)\cdot j^{-5}).\end{aligned}$$

Therefore (13) $= O(f(\lambda_j)\cdot j^{-4})$.

Let's study now the covariance term, $0<k<j-2<N$,

$$E[w^T(\lambda_j)\overline{w^T(\lambda_k)}] = \frac{1}{2\pi\sum h_t^2}\int_{-\pi}^{\pi}D^T(\lambda_j-\lambda)D^T(\lambda-\lambda_k)f(\lambda)d\lambda. \tag{14}$$

The covariance (14) can be expanded as

$$\frac{1}{2\pi\sum h_t^2}\left[\int_{(\lambda_k+\lambda_j)/2}^{2\lambda_j}\{f(\lambda)-f(\lambda_j)\}\,D^T(\lambda_j-\lambda)D^T(\lambda-\lambda_k)d\lambda\right. \tag{15}$$

$$+\int_{\lambda_k/2}^{(\lambda_k+\lambda_j)/2}\{f(\lambda)-f(\lambda_k)\}\,D^T(\lambda_j-\lambda)D^T(\lambda-\lambda_k)d\lambda \tag{16}$$

$$-\int_{\lambda_k/2}^{(\lambda_k+\lambda_j)/2}\{f(\lambda_j)-f(\lambda_k)\}\,D^T(\lambda_j-\lambda)D^T(\lambda-\lambda_k)d\lambda \tag{17}$$

$$\left.+\left\{\int_{2\lambda_j}^{\pi}+\int_{-\pi}^{\lambda_k/2}\right\}\{f(\lambda)-f(\lambda_j)\}\,D^T(\lambda_j-\lambda)D^T(\lambda-\lambda_k)d\lambda\right]. \tag{18}$$

Then, for N and ϵ chosen as before, for the following intervals in (18)

$$\left|\int_{-\pi}^{-\epsilon}+\int_{\epsilon}^{\pi}\right| = O\left(\frac{1}{N^5\epsilon^6}\int_{-\pi}^{\pi}[f(\lambda)+f(\lambda_j)]\,d\lambda\right) = O(f(\lambda_j)\cdot N^{-5}),$$

and

$$\left|\int_{-\epsilon}^{-\lambda_j}+\int_{2\lambda_j}^{\epsilon}\right| = O\left(\left[\max_{\lambda_j\le\lambda\le\epsilon} f(\lambda)+f(\lambda_j)\right] N^{-1}\int_{2\lambda_j}^{\pi}\left|D^T(\lambda_j-\lambda)D^T(\lambda-\lambda_k)\right| d\lambda\right)$$
$$= O\left(f(\lambda_j)\cdot N^{-5}\int_{2\lambda j}^{\infty}\lambda^{-6}d\lambda\right) = O\left(f(\lambda_j)\cdot j^{-5}\right).$$

Now (15) can be bounded by

$$\left|\int_{(\lambda_k+\lambda_j)/2}^{2\lambda_j}\right| = O\left(N^{-1}\sup_{(\lambda_k+\lambda_j)/2\le\lambda\le 2\lambda_j}|f'(\lambda)D^T(\lambda-\lambda_k)|\int_{(\lambda_k+\lambda_j)/2}^{2\lambda_j}|\lambda_j-\lambda||D^T(\lambda_j-\lambda)|d\lambda\right)$$
$$= O\left(N^{-1}\cdot f(\lambda_j)\lambda_j^{-1}\cdot N^{-2}\lambda_k^{-3}\int_0^{\lambda_j}\lambda|D^T(\lambda)|d\lambda\right) = O(f(\lambda_j)\cdot k^{-3}j^{-1}).$$

Next, for $k\ge j/2$, (16) is

$$O\left(N^{-1}\sup_{\lambda_k/2\le\lambda\le(\lambda_k+\lambda_j)/2}|f'(\lambda)D^T(\lambda_j-\lambda)|\int_{\lambda_k/2}^{(\lambda_k+\lambda_j)/2}|\lambda-\lambda_k||D^T(\lambda-\lambda_k)|d\lambda\right)$$
$$= O\left(N^{-1}\cdot f(\lambda_k)\lambda_k^{-1}\cdot N\int_0^{\lambda_j}\lambda|D^T(\lambda)|d\lambda\right) = O(f(\lambda_k)\cdot k^{-1}),$$

and when $k<j/2$, (16) is

$$O\left(N^{-1}\sup_{\lambda_k/2\le\lambda\le(\lambda_k+\lambda_j)/2}|f(\lambda)+f(\lambda_k)|\int_{\lambda_k/2}^{(\lambda_k+\lambda_j)/2}|D^T(\lambda_j-\lambda)D^T(\lambda-\lambda_k)|d\lambda\right)$$
$$= O\left(N^{-1}\cdot f(\lambda_k)\cdot N^{-2}\lambda_{(j-k)}^{-3}\int_0^{\lambda_j}|D^T(\lambda)|d\lambda\right) = O(f(\lambda_k)\cdot j^{-3}).$$

because $\int_0^{2\lambda_j}\lambda|D^T(\lambda)|d\lambda = O(N^{-1})$. For $k\ge j/2$, (17) is

$$O\left(N^{-1}(\lambda_j-\lambda_k)\sup_{\lambda_k\le\lambda\le\lambda_j}|f'(\lambda)|\int_{\lambda_k/2}^{(\lambda_k+\lambda_j)/2}|D^T(\lambda_j-\lambda)||D^T(\lambda-\lambda_k)|d\lambda\right)$$
$$= O\left(N^{-1}\cdot f(\lambda_k)\cdot\lambda_k^{-1}\int_0^{\lambda_j}|D^T(\lambda)|d\lambda\right) = O(f(\lambda_k)\cdot k^{-1}),$$

and when $k<j/2$

$$O\left(N^{-1}\sup_{\lambda_k/2\le\lambda\le(\lambda_k+\lambda_j)/2}|f(\lambda)+f(\lambda_k)|\int_{\lambda_k/2}^{(\lambda_k+\lambda_j)/2}|D^T(\lambda_j-\lambda)D^T(\lambda-\lambda_k)|d\lambda\right)$$
$$= O\left(N^{-1}\cdot f(\lambda_k)\cdot N^{-2}\lambda_{j-k}^{-3}\int_0^{\lambda_j}|D^T(\lambda)|d\lambda\right) = O(f(\lambda_k)\cdot j^{-3}).$$

Finally,

$$\left|\int_{-\lambda k/2}^{\lambda_k/2}\right| = O\left(N^{-1}\max_{-\lambda_k/2\le\lambda\le\lambda_k/2}\left|D^T(\lambda-\lambda_k)D^T(\lambda_j-\lambda)\right|\int_{-\lambda_k/2}^{\lambda_k/2}[f(\lambda)+f(\lambda_j)]\,d\lambda\right)$$
$$= O\left(N^{-5}\lambda_k^{-2}\lambda_j^{-3}f(\lambda_k)\right) = O(f(\lambda_k)\cdot k^{-2}j^{-3}).$$

Then the bound for (13) is $O([f(\lambda_k)f(\lambda_j)]^{1/2}\cdot k^{-1})$.

A similar bound for the term corresponding to (d),

$$E[w^T(\lambda_j)w^T(\lambda_k)] = \frac{1}{2\pi\sum h_t^2}\int_{-\pi}^{\pi}D^T(\lambda_j-\lambda)D^T(\lambda+\lambda_k)f(\lambda)d\lambda,$$

can be obtained easily using the same methods as for (c), since we do not need to distinguish between frequencies j and k too close.

9 Appendix: Proof of Theorem 3

We do that in three steps.

First, we argue that all the previous results concerning the asymptotic distribution of $\overline{I}_{\epsilon j}$, $j \in \{j_{(1)}, j_{(2)}, \ldots, j_{(k)}\}$, still go through for the tapered version $\overline{I}_{\epsilon j}^{T}$ if $j_{(1)} + 2 < j_{(2)}, \ldots, j_{(k-1)} + 2 < j_{(k)}$.

The reason is the following: Chen and Hannan (1980) results are based on the exact uncorrelatedness of the discrete Fourier transform of the i.i.d. sequence of ϵ_t at different Fourier frequencies, so the periodograms ordinates are approximately independent. Therefore, the real and imaginary components the tapered Fourier transforms of ϵ_t are still exactly uncorrelated, if we consider only one periodogram ordinate of every three, as we did in the definition of $\widehat{d}^T$. Then an equivalent Edgeworth expansion for the density of the vector of real and imaginary components of $\overline{I}_{\epsilon k}^{T}$ is valid as before, since each of the tapered Fourier transforms is a (fixed) linear combination of three Fourier transforms with valid Edgeworth expansions for their densities.

Second. Therefore, corresponding to Lemmas 2 to 4 in the non tapered case, we obtain under the conditions of the theorem and $J \geq 2$, $k > 1$ and the same definitions as before, but now with the tapered periodogram,

$$\overline{I}_{\epsilon k}^{T} > 0 \qquad \text{w.p.1.,}$$

$$\log(1 + F_k^T) = O_P(k^{-1}), \qquad k = \ell + 3J, \ell + 6J, \ldots, m,$$

and reasoning as for expression (3.17) in Robinson (1995b), using now part (a) of our Theorem 2, instead of Theorem 2 of Robinson (1995a), $\alpha > 1$,

$$\log\left(1 + \frac{\delta_k^T}{H_k^T}\right) = O_P\left(k^{-\alpha/2}\right), \qquad k = \ell + 3J, \ell + 6J, \ldots, m.$$

Now, from the proof of Theorem 1 we obtain, $\alpha < 2$,

$$\begin{aligned}
\widehat{d}^T &= d + \left(\sum_k \Lambda_k^2\right)^{-1}\left(\sum_k \Lambda_k \log 2\pi \overline{I}_{\epsilon k}^{T}\right) + O_P\left(\frac{\log N}{m}\sum_k k^{-\alpha/2}\right) + O([mN^{-1}]^{\alpha}) \\
&= d + \xi_N^T + O_P\left(m^{-\alpha/2}\log N + m^{-1}\log N \log m + [mN^{-1}]^{\alpha}\right), \quad \text{say,} \\
&= d + \xi_N^T + o_P(m^{-1/2}),
\end{aligned}$$

using Assumption 7. Hence, the asymptotic distribution of $m^{1/2}(\widehat{d}^T - d)$ can be approximated by that of $m^{1/2}\xi_N^T$.

Third. We denote by $\xi_N^{T,*}$ and $\log 2\pi \overline{I}_{\epsilon,k}^{T,*}$ the corresponding random variables when the ϵ_t are Gaussian. Now we follow related arguments to those of Robinson (1995a): we will show that the moments of all orders of $m^{1/2}\xi_N^T$ converge to those of $m^{1/2}\xi_N^{T,*}$ which are bounded, from Robinson's (1995a) Theorem 3's proof. Next, the uncorrelatedness of the real and imaginary components of $\overline{I}_{\epsilon,k}^{T,*}$ for different frequencies implies the independence and equal distribution (due to the Gaussianity) of $\overline{I}_{\epsilon,k}^{T,*}$ and of $\log \overline{I}_{\epsilon,k}^{T,*}$ at different frequencies. Therefore, $m^{1/2}\xi_N^{T,*}$ is a sum of i.i.d. variables with bounded moments and by the Lindeberg-Feller CLT it is asymptotically normal (whose first two moments can be obtained from the proof of Theorem 1). Because each moment of the variate $m^{1/2}\xi_N^{T,*}$ is bounded uniformly in N, all these moments converge to those of the corresponding normal distribution. Hence, as all moments of $m^{1/2}\xi_N^T$ converge to those of $m^{1/2}\xi_N^{T,*}$, $m^{1/2}\xi_N^T$ is easily found asymptotically normal distributed by the method of moments.

Therefore, it only remains to prove that the moments of all orders of $m^{1/2}\xi_N^T$ converge to those of $m^{1/2}\xi_N^{T,*}$, so there is not influence from the higher order cumulants of ϵ_t.

Lemma 6 *Under the assumptions of Theorem 3, the moments of all orders of $m^{1/2}\xi_N^T$ converge to those of $m^{1/2}\xi_N^{T,*}$ as $N \to \infty$.*

Proof of Lemma 6 In order to make our arguments clearer we are going to consider in an initial stage the non-tapered and non-pooled ($J = 1$) case. After this, we will show that the same conclusions apply for the tapered case for any $J > 1$.

Following the arguments of the second part in the proof of Theorem 1, it is easy to check that the moments and cross moments of all orders of $\log 2\pi \overline{I}_{\epsilon,k}^T$, $k = \ell + 3J, \ldots, m$, converge to those we would obtain if the ϵ_t's were actually Gaussian, with error $O(N^{-1})$, since $\int_0^\infty (\log x)^a (1 + x^4)^{-1} dx < \infty$ for all $a \geq 0$.

However, this result is not enough to approximate the moments of $m^{1/2}\xi_N^T$, which is an (infinite) weighted sum of the $\log 2\pi \overline{I}_{\epsilon,k}^T$. When ϵ_t has bounded moments of all orders, we can obtain an Edgeworth expansion for the density of the Fourier transform of ϵ_t of any order s fixed, under the same assumptions of Lemma 5. In Chen and Hannan (1980) the second term P_2 is presented, although it is not totally correct in their notation. The exact shape of these higher order terms in a general Edgeworth approximation is fundamental for our proof and we dedicate some space to that.

Edgeworth approximation. The validity of an Edgeworth approximation for the real and imaginary components of the discrete Fourier transform of ϵ_t of any order $s > 1$, when enough moments exist, follows from Chen and Hannan's (1980) Lemma 2, since their proof generalizes immediately for any order of approximation, not just 2. For any $s = 0, 1, \ldots,$ fixed, we can obtain that the vector W_N (see Lemma 5) has density q_N for all sufficiently large N and

$$\sup_{\mathbf{y} \in \mathcal{R}^{2k}} (1 + \|\mathbf{y}\|^s) \left| q_N(\mathbf{y}) - \sum_{r=0}^{s} N^{-r/2} P_r(-\phi : \overline{\chi}_{\nu,N})(\mathbf{y}) \right| = O(N^{-(s+1)/2}), \tag{19}$$

where P_r are polynomials with coefficients depending on the joint cumulants of Y_t, $\overline{\chi}_{\nu,N}$, multiplied by the $2k$th standard multivariate Normal density ϕ (given the covariance structure of the discrete Fourier transforms). Following Bhattacharya and Rao (1976), we find that $P_r(-\phi : \overline{\chi}_{\nu,N}) = \tilde{P}_r(-D : \overline{\chi}_{\nu,N})\phi$, where, for nonnegative integer vectors $\nu = (\nu(1), \nu(2), \ldots, \nu(2k))$ of $2k$ dimensions,

$$D^{\nu} = \left(\frac{\partial}{\partial y_1}\right)^{\nu(1)} \cdots \left(\frac{\partial}{\partial y_{2k}}\right)^{\nu(2k)},$$

with

$$\begin{aligned}\tilde{P}_r(z:\overline{\chi}_{\nu,N}) &= \sum_{n=1}^{r}\frac{1}{n!}\left\{\sum_{j_1,\dots j_n}{}^{*}\frac{\chi_{j_1}(z)}{j_1!}\frac{\chi_{j_2}(z)}{j_2!}\cdots\frac{\chi_{j_n}(z)}{j_n!}\right\}\\ &= \sum_{n=1}^{r}\frac{1}{n!}\left\{\sum_{j_1,\dots j_n}{}^{*}\left(\sum{}^{**}\frac{\chi_{\nu_1}\cdots\chi_{\nu_n}}{\nu_1!\cdots\nu_n!}z^{\nu_1+\cdots+\nu_n}\right)\right\}.\end{aligned}\tag{20}$$

The summation $\sum^{*}$ is over all n-tuples of positive integers $(j_1,\dots,j_n)$ satisfying

$$\sum_{i=1}^{n} j_i = r, \qquad j_i = 1,2,\dots,r \qquad (1\le i\le n),\tag{21}$$

and the $\sum^{**}$ denotes summation over all n-tuples of nonnegative integral vectors $(\nu_1,\dots,\nu_n)$ satisfying

$$|\nu_i| = j_i + 2 \qquad\qquad (1\le i\le n),$$

where we have been using the usual multivariate notation, $|\nu_i| = \sum_{j=1}^{2k}\nu_i(j)$ (see Bhattacharya and Rao (1976) for details). In particular, $\tilde{P}_0\equiv 1$

$$\begin{aligned}\tilde{P}_1(z:\overline{\chi}_{\nu,N}) &= \frac{\chi_3(z)}{3!} = \sum_{|\nu|=3}\frac{\overline{\chi}_{\nu,N}}{\nu!}z^{\nu},\\ \tilde{P}_2(z:\overline{\chi}_{\nu,N}) &= \frac{\chi_4(z)}{4!}+\frac{1}{2!}\left(\frac{\chi_3(z)}{3!}\right)^2,\\ \tilde{P}_3(z:\overline{\chi}_{\nu,N}) &= \frac{\chi_5(z)}{5!}+\frac{\chi_4(z)\chi_3(z)}{3!\,4!}+\frac{1}{3!}\left(\frac{\chi_3(z)}{3!}\right)^3,\end{aligned}$$

and in general

$$\chi_r(z) = \sum_{|\nu|=r}\frac{\overline{\chi}_{\nu,N}}{\nu!}z^{\nu},$$

where $\nu_i! = \nu_i(1)!\cdots\nu_i(2k)!$, $0! = 1$. Then we can see that $P_r(\mathbf{y})$ is a polynomial in the components of $\mathbf{y}$ (times ϕ), with coefficients which are functions of the joint cumulants of Y_t of order ν (i.e. of the components in the vector Y_t with exponent in ν different form zero), $\overline{\chi}_{\nu,N}$, and of the Hermite polynomials of order ν, $H_\nu(\mathbf{y})$, obtained from (the derivatives of) $\phi(\mathbf{y})$. Following our previous discussion in the proof of Theorem 1 and the comments above, we stress some properties that we will use later:

- Using expansion (19), the first term of the expectations of functions of the periodogram of ϵ_t is always exactly equal to the Gaussian expectation, so we need only to concentrate on the higher order terms of an approximation of the sufficient order (to be determined later).

- When $|\nu|$ is odd, the polynomial function $H_\nu(\mathbf{y})$ will be odd in at least one of the components of $\mathbf{y}$. Then, all the summands in P_r with r odd will be also odd in at least one of the components of $\mathbf{y}$. As we are going to consider the expectation of even functions of $\mathbf{y}$, (i.e. logarithm of the periodogram minus a constant) we need to consider only terms P_r with r even ($r = 0,2,\dots$).

- Furthermore, the cumulants $\overline{\chi}_{\nu,N}$ will be exactly zero in many situations (i.e., for many vectors ν), due to the special nature of the vector W_N. In other cases these cumulants will be different from zero only under linear restrictions on some of the particular frequencies $(\lambda_{j(1)}, \ldots, \lambda_{j2k)})$ of the periodogram ordinates that we are considering in each moment.

Moments. Since $\sum_k \Delta_k \equiv 0$, in contrast with the proof of Theorem 1, we substitute now in the definition of ξ_N the actual mean of $\log 2\pi I_{\epsilon,k}$ by the mean it would have in the Gaussian case, $\psi(J)$, obtaining (without need to make explicit the $\log 2$ adjustment), with $J = 1$,

$$\xi_N = \left(\sum_k \Lambda_k^2\right)^{-1} \sum_k \Lambda_k \left(\log 2\pi I_{\epsilon k} - \psi(1)\right).$$

Denote by E_s and E_s^* the sth moments of $m^{1/2}\xi_N^T$ and $m^{1/2}\xi_N^{T,*}$, respectively. Then, for $s = 3, 4, \ldots$,

$$\begin{aligned} E_s &\equiv E\left[\left(m^{1/2}\left(\sum \Lambda_k^2\right)^{-1} \sum_k \Lambda_k \left(\log 2\pi I_{\epsilon k} - \psi(1)\right)\right)^s\right] \\ &= m^{s/2}\left(\sum \Lambda_k^2\right)^{-s} \sum_{j=1}^{s} \sum_p c_p \sum_{k(1)} \sum_{k(2)\neq} \cdots \sum_{k(j)\neq} \Lambda_{k(1)}^{p(1)} \Lambda_{k(2)}^{p(2)} \cdots \Lambda_{k(j)}^{p(j)} \\ &\quad \times E\left[\left(\log I_{\epsilon,k(1)}^T - \psi(1)\right)^{p(1)} \cdots \left(\log I_{\epsilon,k(j)}^T - \psi(1)\right)^{p(j)}\right], \end{aligned} \tag{22}$$

where the index $k(i) \neq$ means that the summation is for all the values of $k(i) \neq k(1), \ldots, k(i-1)$ (so we only make explicit frequencies that are always different) and the sum in p is for all vectors of positive integers $(p(1), \ldots, p(j))$ such that $\sum_{i=1}^{j} p(i) = s$, and c_p is a number that depends only on p. Obviously, when $j = s$, all $p(i) = 1$ and when $j = 1$, $p(1) = s$. Recall also that $(\sum \Lambda_k^2)^{-1} = O(m^{-1})$.

Now, the idea is to substitute for N big enough each of the expectations in (22) by an integral over $\mathcal{R}^{2j}$, approximating the true probability density of the vector of periodogram ordinates by a $2j$-dimensional Edgeworth expansion of the form (19). As we have commented the first term (in N^0) of the Edgeworth approximation always gives the corresponding Gaussian expectation E_s^*:

$$\begin{aligned} E_s - E_s^* &= m^{s/2}\left(\sum \Lambda_k^2\right)^{-s} \sum_{j=1}^{s} \sum_p c_p \sum_{k(1)} \sum_{k(2)\neq} \cdots \sum_{k(j)\neq} \Lambda_{k(1)}^{p(1)} \Lambda_{k(2)}^{p(2)} \cdots \Lambda_{k(j)}^{p(j)} \\ &\quad \times \int_{\mathcal{R}^{2j}} \left\{\sum_{r=2}^{r^{\max}} P_r(\mathbf{y}) N^{-r/2} + O(N^{-(r^{\max}+1)/2})(1 + \|\mathbf{y}\|^4)^{-1}\right\} \\ &\quad \times \left(\log(y_{a,k(1)}^2 + y_{b,k(1)}^2) - \psi(1)\right)^{p(1)} \cdots \left(\log(y_{a,k(j)}^2 + y_{b,k(j)}^2) - \psi(1)\right)^{p(j)} d\mathbf{y}, \end{aligned} \tag{23}$$

(with $r_{\max} \leq s$ to be determined later). Then we need to check that the contribution from all higher order terms with $r \geq 2$ in (23) is negligible. Now for s fixed, we study all the terms in (23) with different values of j.

Consider first the terms in (23) for which $\underline{j \leq 1 + s/2}$. Using (10) and that $\int_0^\infty (\log x)^b (1+x^4)^{-1} dx < \infty$, $b \geq 0$, the contribution to $E_s - E_s^*$ of each of the higher order terms P_r, $r > 0$, is $O\left(m^{-s/2+j} N^{-1}\right) = O(mN^{-1}) = o(1)$, just using the order of magnitude of the error term of an Edgeworth approximation with only P_0 and P_2, since the term P_1 cancel out.

Therefore, we need only to consider terms where $j > 1 + s/2$. The main idea to deal with these terms is the following. Since we have $j > 1 + s/2$ summands, there should be some of them, h say, with exponent $p(i) = 1$. In fact

$$h \geq 2j - s \geq 3.$$

Then we know that whenever $h \geq 1$ the leading term in the approximation for the correspondent expectation (the Gaussian part) is exactly zero (i.e. $E_s^* = 0$), given the uncorrelatedness of the discrete Fourier transform at Fourier frequencies (even in the non-Gaussian case). We will show that this orthogonality property of the first (Gaussian) term is transferred to some extend to the higher order terms. The reason is that for each periodogram ordinate (i.e. for each couple of variables in $\mathbf{y}$), some of the contributions from the higher order terms in (19) are still the Gaussian ones given by $\phi(y_a)\phi(y_b)$ (i.e., we have not taken derivatives w.r.t. those variables), given null contribution for the whole expectation when this periodogram ordinate has power $p(i) = 1$. [The same argument can be used for any exponent $p(i)$ odd, but we will not need it].

We illustrate this idea with an example. Consider P_2, with

$$\begin{aligned}
\tilde{P}_2(z : \overline{X}_{\nu,N}) &= \sum_{|\nu|=4} \frac{\overline{X}_{\nu,N}}{\nu!} z^{\nu} + \frac{1}{2}\left(\sum_{|\nu|=3} \frac{\overline{X}_{\nu,N}}{\nu!} z^{\nu}\right)^2, \\
&= \sum_{|\nu|=4} \frac{\overline{X}_{\nu,N}}{\nu!} z^{\nu} + \frac{1}{2}\sum_{|\nu|=3}\left(\frac{\overline{X}_{\nu,N}}{\nu!}\right)^2 z^{2\nu} + \frac{1}{2}\sum_{|\nu|=3}\ \sum_{|\nu'|=3,\nu\neq\nu'} \frac{\overline{X}_{\nu,N}}{\nu!}\frac{\overline{X}_{\nu',N}}{\nu'!} z^{\nu+\nu'}.
\end{aligned}$$

When we substitute z by $-D$, to obtain P_2, we observe that in each of the terms of the last expression we take at most 4, 3 or 6 derivatives, respectively, with respect to the vector of $2j$ components $\mathbf{y}$. Therefore, all but at most 4, 3 or 6 functions $\phi(y_i)$ in $\phi \equiv \phi(y_1)\cdots\phi(y_{2j})$ are not affected by the differential operator.

Then, for each periodogram ordinate, with frequency $\lambda_{j(i)}$, and each of the terms in P_2 (with ν, 2ν or ν and ν') we can obtain the following:

- If the periodogram has exponent $p(i) = 1$ and neither of its two components in $\mathbf{y}$ is included in ν (or in ν'), then we have in (23) an integral of the form

$$\int_{\mathcal{R}}\int_{\mathcal{R}} \left[\log(y_a^2 + y_b^2) - \psi(1)\right]\phi(y_a)\phi(y_b)dy_a dy_b \equiv 0, \tag{24}$$

  and this whole term does not contribute, since the whole integral to approximate this term of E_s is zero because the variables y_a and y_b do not appear in any other factor of this particular summand (23).

- For any exponent $p(i)$, if any component of the vector $\mathbf{y}$ is included in ν (or ν') with odd coordinate, then the term in P_2 will be odd in that variable y_a, say, and again the contribution of these terms is null, since the periodogram is even in its real and imaginary components y_a and y_b.

In conclusion, *we need at least two derivatives with respect to one of the two variables which have* $p(i) = 1$, or in other terms, *only the terms in* P_r *which consider Hermite polynomials with an even number in its order vector corresponding to one of the two variables with* $p(i) = 1$ *will have contribution different from zero.*

In the particular case of P_2 we only need to consider the following generic vectors ν. When $|\nu| = 4$, only those vectors ν with coordinates 0, 2 or 4. For the $|\nu| = 3$ terms, any combination is valid from this point of view, since all the terms are squared, and for the $|\nu|$, $|\nu'| = 3$ terms, coordinates 1 or 3 are allowed in ν only if the coincide with another coordinate 1 or 3 in ν', in order to take always an even number of derivatives w.r.t. to any of the variables y_i. However, the number of such terms is limited by the number r.

Then with $|\nu| = 4$, the maximum number of frequencies *affected* by the derivative, MNFA say [in the sense that we are taking an even number of derivatives w.r.t. any of the components of the periodogram at this particular frequency] is 2 and with $|\nu| = 3$ and/or $|\nu'| = 3$ this number is 3.

Consider the worst case (with the biggest number of summations), where $j = s$, so all the $p(i) = 1$. Then for P_2, the contribution when $s > 3$ is zero, since there will be always at least one integral equal to zero by (24), as none of its components is included in the differentiation (i.e. there are at least 4 possible orthogonal conditions like (24), and only three can be destroyed by the differentiation of ϕ). When $s = 3$ and $j = 3$ we obtain that any term will contribute $O(m^{3/2}N^{-1})$.

For example when $j = s - 1$, so there are at least $s - 2$ exponents $p(i) = 1$, and we have that for $s > 5$ the contribution of P_2 is zero, for $s = 3$, $O(m^{1/2}N^{-1})$, for $s = 4$, $O(mN^{-1})$ and for $s = 5$, $O(m^{3/2}N^{-1})$.

In general, for any $j > 1 + s/2$, (the same s fixed) since $h \geq 2j - s$ there are only terms in P_2 that contribute to E_s if

$$\min h = 2j - s \leq 3 = \text{MNFA},$$

and in this case their contribution is of order $O(m^{j-s/2}N^{-1}) = O(m^{3/2}N^{-1})$. We can check that related bounds hold for any term P_r. Although it is possible to assume $m^{3/2}N^{-1} \to 0$ as N increases, in order to make all these bounds $o(1)$ as we need, there is another point that will allow us to obtain the same results with just $mN^{-1}(\log N)^c \to 0$, any finite $c > 0$, implied by the assumptions of the Lemma.

Cumulants. The bounds above have been constructed for P_2 considering that 3 frequencies were affected in the term corresponding to cumulants with $|\nu| = 3$. The question is when are these cumulants different from zero. For any three frequencies $\lambda_{j_1}, \lambda_{j_2}, \lambda_{j_3}$, (possibly repeated) the cumulant $\chi_{\nu,N}$, $|\nu| = 3$, is of any of the following four types, with κ_3 being the third cumulant of ϵ_t:

$$(i) \qquad \frac{\kappa_3}{N} \sum_{t=1}^{N} \cos t\lambda_{j_1} \cos t\lambda_{j_2} \cos t\lambda_{j_3},$$

$$(ii) \qquad \frac{\kappa_3}{N}\sum_{t=1}^{N}\cos t\lambda_{j_1}\cos t\lambda_{j_2}\sin t\lambda_{j_3},$$

$$(iii) \qquad \frac{\kappa_3}{N}\sum_{t=1}^{N}\cos t\lambda_{j_1}\sin t\lambda_{j_2}\sin t\lambda_{j_3},$$

$$(iv) \qquad \frac{\kappa_3}{N}\sum_{t=1}^{N}\sin t\lambda_{j_1}\sin t\lambda_{j_2}\sin t\lambda_{j_3}.$$

Now using $\sin\lambda = (e^{i\lambda} - e^{-i\lambda})/(2i)$, $\cos\lambda = (e^{i\lambda} + e^{-i\lambda})/2$, and the orthogonality of Dirichlet kernel at Fourier frequencies, we can see that **all** the cumulants $\chi_{\nu,N}$ with $|\nu| = 3$ will only be different from zero if there is a linear restriction between the frequencies $\lambda_{j_1}, \lambda_{j_2}, \lambda_{j_3}$. [The same holds for any odd-order $|\nu|$ cumulant: we need a linear restriction between the frequencies to make the sum in t different from zero. However, some even-order cumulants can be different from zero without restrictions, due to symmetries]. Then, all the bounds have to be multiplied by $m^{-1}\log N$, since one of the summations $\sum_k$ in (23) has now been cancelled out due to the linear restriction with the other (two) summation(s), and $\sup|\Lambda_k| = O(\log N)$. Finally, we obtain a contribution of $O(m^{1/2}N^{-1}\log N) = o(1)$, for any term with $|\nu| = 3$.

If we consider the case of the cumulants with $|\nu| = 4$, here at most 2 frequencies are affected by the differentiation, so they contribute to E_s only if

$$\min h = 2j - s \leq 2 = \text{MNFA}.$$

Then, we obtain a bound for their contribution of $O(m^{j-s/2}N^{-1}) = O(mN^{-1}) = o(1)$ (as no restrictions are required in this case to make some cumulants in $\chi_{\nu,N}$ different from zero).

Let's study now the contribution from a generic polynomial P_r, $r \geq 4$. We only need to consider expansions up to $r \leq r_{\max} = 2[\frac{s-1}{2}]$, (where $[\cdot]$ means integer part) since the bound in (23) due to the error term in the Edgeworth expansion with $P_{2[\frac{s-1}{2}]}$ it is immediately $o(1)$ from the exponent $N^{-1-[\frac{s-1}{2}]}$ in it and the boundedness of the corresponding integral.

Now from (20), the different terms in P_r will include terms with combinations of cumulants

$$\kappa_{r+2},\ \kappa_{r+1}\kappa_3,\ \kappa_r\kappa_4,\ \ldots,\ (\kappa_4)^{r/2},\ \ldots,\ (\kappa_3)^r,$$

corresponding to all possible combinations of frequencies inj the vector $\mathbf{y}$.

We will only need to consider combinations of cumulants of the form $(\kappa_4)^{(r-a)/2}(\kappa_3)^a$, for even a, $0 \leq a \leq r$ when the MNFA is now $r + a/2$, which requires at least $a/2$ restrictions. The reason is that with $(\kappa_4)^{r/2}$ we maximize the number of frequencies affected without any restrictions, and on the other hand, with $(\kappa_3)^r$ we maximize the number of frequencies affected, in general, with and without restrictions. We show bellow that any other combination of higher order cumulants will always provide a smaller MNFA or more restrictions that the combinations between cumulants of order three and four, which optimize the number of degrees of freedom when consider the destruction of orthogonality restrictions due to equation (24).

Denoting by NRES the (minimum) number of linear restrictions necessary to make the cumulants considered different from zero, the contribution to (23) of these terms is of order, for fixed s, $j > 1+s/2$,

$$\begin{aligned}
& m^{-s/2}\sum_j^s m^j \sum_r N^{-r/2} I\{2j-s \leq \text{MNFA}\}(m^{-1}\log m)^{\text{NRES}} \\
= \;& O\left(\max_a m^{-s/2}\sum_j^s m^j \sum_r N^{-r/2} I\{2j-s \leq r+a/2\}(m^{-1}\log m)^{a/2}\right) \\
= \;& O\left(\max_{a,j} m^{j-s/2}\sum_r N^{-r/2} I\{2j-s \leq r+a/2\}(m^{-1}\log m)^{a/2}\right) \\
= \;& O\Big(\max_r m^{a/4+r/2} N^{-r/2}(m^{-1}\log m)^{a/2}\Big) \\
= \;& O\Big(\max_r m^{r/2}N^{-r/2}\Big) = o(1),
\end{aligned}$$

with $a = 0$, so the contribution is always negligible.

Let us show that we do not need to consider other set of cumulants formally. Consider the case with maximum number of frequencies affected, without restrictions: $(\kappa_4)^{r/2}$, so we have the typical term with contribution of the biggest order of magnitude without restrictions. Then we study the question: Can the introduction of $b \geq 1$ restrictions generate terms of bigger order of magnitude in P_r than the one corresponding to $(\kappa_4)^{r/2}$ (for any j, and s fixed)?

Seeking the least favourable situation, the new c restrictions will be used to maximize the number of frequencies affected by the differentiation, substituting certain number of powers of κ_4 with a generic term in the odd-order cumulants (to take advantage of the restrictions) like

$$(\kappa_{c_1})^2(\kappa_{c_2})^2\cdots(\kappa_{c_b})^2,$$

where the $c_i \geq 3$ are odd, possibly equal. This will increase MNFA by $\sum_i c_i$, and the reduction in the exponent of κ_4, in order to satisfy (21), is of magnitude

$$\sum_i (c_i - 2).$$

This reduction will lower MNFA (by the contribution of κ_4) in $2\sum_i(c_i-2)$ units. The global effect on MNFA is finally

$$\sum_i c_i - 2\sum_i (c_i-2) = 4b - \sum_i c_i$$

Therefore in a generic bound (for any j) for the contribution of these terms, $O(m^{j-s/2}m^{-\text{NRES}}N^{-r/2}) = O(m^{\text{MNFA}/2}m^{-\text{NRES}}N^{-r/2})$ [by $2j-s \leq \text{MNFA}$], the global effect of introducing the new b restrictions is of $O((\log N)^b)$ times $O(m)$ to the power of

$$\frac{4b-\sum_i c_i}{2} - b = b - \frac{1}{2}\sum_i c_i.$$

Since $c_i \geq 3$ we have that the effect is [forgetting about the logarithm term] at most of order $O(m^{b-3b/2}) = O(m^{-b/2}) = O(1)$, which is exactly what we had obtained previously, the term with biggest contribution is that with $(\kappa_4)^{r/2}$, confirming that the relevant cases are the ones with lower-order cumulants, with or without restrictions (κ_3 and κ_4).

The last two points which need justification are the tapering and the pooling:

- **Tapering**: as we have commented before, an equivalent Edgeworth expansion for the real and imaginary components of tapered Fourier transform of ϵ_t is valid, since they are fixed linear combinations of the components of the usual Fourier transform. Also, because we are considering frequencies that are $3J\lambda_1$ apart, at least, we guarantee the uncorrelatedness of the different variables in $\mathbf{y}$. In this way the Edgeworth expansion is based again in the standard Normal density, so the differentiation process is performed separately for each variable in $\mathbf{y}$.

  Furthermore, the comments about the restrictions to obtain odd-order cumulants different from zero apply equally in the tapered case, since for the frequencies considered, the orthogonality properties of D^T (and of its real and imaginary parts) are the same as for the Dirichlet kernel D.

- **Pooling**: the difference now is that each pooled periodogram (tapered or not) depends on $2J$ components of the basic vector $\mathbf{y}$ instead on just 2 (single periodogram) as before. This will not affect any of the results, since we have only used the fact that in each summand of (23) there are j different $\log \bar{I}_k$ functions, but not that the vector of variables $\mathbf{y}$ (in the Edgeworth expansion required to approximate each expectation) were of dimension $2j$ ($2jJ$ now). The same comments about the differentiation to obtain the Hermite polynomials and the cancellation of integrals go through here again, as we have considered the cases where just differentiation (an even number of times) w.r.t. to one single component of the periodogram destroys the orthogonality condition (24).

Then, the proof of the Theorem is complete.

References

[1] Beran, J. (1994) *Statistics for Long Memory Processes.* London: Chapman Hall.

[2] Bhattacharya, R.N. and R.R. Rao (1976) *Normal Approximation and Asymptotic Expansions.* New York: Wiley.

[3] Bloomfield, P. (1976) *Fourier Analysis of Time Series: An Introduction.* New York: Wiley.

[4] Chen, Z.-G. and E.J. Hannan (1980) The distribution of periodogram ordinates. *Journal of Time Series Analysis* 1, 73-82.

[5] Comte, F. and C. Hardouin (1995a) Regression on log-regularized periodogram under assumption on bounded spectral densities: the non fractional and the fractional cases. Working Paper *CREST*.

[6] Comte, F. and C. Hardouin (1995b) Regression on log-regularized periodogram for fractional models at low frequencies. Working Paper *CREST*.

[7] Dahlhaus, R. (1988) Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Annals of Statistics* 16, 808-841.

[8] Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Annals of Statistics* 17, 1749-1766.

[9] Feller, W. (1971) *An Introduction to Probability Theory and its Applications,* Vol. 2, 2nd edition. New York: Wiley.

[10] Fox, R. and M.S. Taqqu (1989) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* 14, 517-532.

[11] Geweke, J. and S. Porter-Hudak (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221-238.

[12] Götze, F. and C. Hipp (1983) Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 64, 211-239.

[13] Gradshteyn, I.S. and I.M. Ryzhik (1980) *Table of Integrals, Series and Products.* Orlando: Academic Press.

[14] Hannan, E.J. (1970) *Multiple Time Series.* New York: Wiley.

[15] Hannan, E.J. and D.F. Nicholls (1977) The estimation of the prediction error variance. *Journal of the American Statistical Association* 72, 834-840.

[16] Hurvich, C.M. and B.K. Ray (1995) Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis* 16, 17-42.

[17] Janas, D. and R. von Sachs (1993) Consistency for non-linear functions of the periodogram of tapered data. *Journal of Time Series Analysis* 16, 585-606.

[18] Percival, D.B. and A.T. Walden (1993) *Spectral Analysis for Physical Applications.* Cambridge: Cambridge University Press.

[19] Lobato, I. and P.M. Robinson (1996) Averaged periodogram estimation of long memory. *Journal of Econometrics* 73, 303-324.

[20] Robinson, P.M. (1986) On the errors-in-variables problem for time series. *Journal of Multivariate Analysis* 19, 240-250.

[21] Robinson, P.M. (1994a) Semiparametric analysis of long-memory time series. *Annals of Statistics* 22, 515-539.

[22] Robinson, P.M. (1994b) Rates of convergence and optimal spectral bandwidth for long range dependence. *Probability Theory and Related fields* 99, 443-473.

[23] Robinson, P.M. (1994c) Time series with strong dependence. In *Advances in Econometrics: Sixth World Congress,* Vol. 1, C.A. Sims, ed., 47-95. Cambridge University Press, Cambridge.

[24] Robinson, P.M. (1995a) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23, 1048-1072.

[25] Robinson, P.M. (1995b) Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23, 1630-1661.

[26] von Sachs, R. (1994) Estimating non-linear functions of the spectral density, using a data-taper. *Annals of the Institute of Statistical Mathematics* 46, 453-474.

[27] Taniguchi, M. (1979) On estimation of parameters of Gaussian stationary processes. *Journal of Applied Probability* 16, 575-591.

[28] Tukey, J.W. (1967) An introduction to the calculation of numerical spectrum analysis. In *Advanced Seminar on Spectral Analysis of Time Series*, ed. B. Harris, pp. 25-46. New York: Wiley.

[29] Velasco, C. (1997) *Higher Order Asymptotic Theory for Nonparametric Time Series Analysis and Related Contributions.* Unpublished PhD Thesis, University of London.